Cognitive behavioral therapy for insomnia: A meta-analysis of long-term effects in controlled studies

Tanja van der Zweerde, Lampros Bisdounis, Simon D. Kyle, Jaap Lancee, Annemieke van Straten

Department of Clinical Psychology, Amsterdam Public Health, VU University, Amsterdam, the Netherlands
Department of Clinical Psychology, University of Amsterdam, Amsterdam, the Netherlands
Sleep and Circadian Neuroscience Institute, Nuffield Department of Clinical Neurosciences, University of Oxford, United Kingdom
PsyQ Amsterdam, Amsterdam, the Netherlands

Article history:
Received 20 November 2018
Received in revised form 25 April 2019
Accepted 7 August 2019
Available online 12 August 2019

Keywords:
Behavior therapy
cognitive therapy
cognitive behavior therapy
long-term
randomized controlled trials
insomnia
sleep initiation or maintenance disorder

Introduction

Insomnia constitutes a serious and common mental health problem with a prevalence of around 6% in the general population (i.e., DSM-IV-TR diagnosis [1]). Furthermore, 30% of the general population suffers from symptoms of insomnia without meeting the criteria of a diagnosis [2]. Insomnia disorder is characterized by a persistent difficulty initiating or maintaining sleep, for three months or longer and for at least three nights a week, resulting in impaired daytime functioning and significant distress [3]. The disorder is associated with high (societal) costs [4] and affects daily life in various domains, such as fatigue, mood changes, declined cognitive ability, physical wellbeing, social relationships and daily tasks [5]. Untreated insomnia often persists for many years [6].

Cognitive behavioral therapy (CBT) is an effective treatment for insomnia. Several systematic reviews and meta-analyses have reported moderate to large short-term post-test effects of CBT for insomnia (CBT-I) [7–9]. These results include robust improvement in insomnia severity (g = 0.98; [9]) and sleep efficiency (g = 0.71 [9]; g = 0.91 [8]; and 9.9% increase in [7]).

In the short term, CBT-I is as effective as pharmacotherapy [10]. However, CBT-I is the preferred treatment according to recommendations in European and American guidelines [11,12]. One reason for advocating CBT-I as first line treatment for insomnia is the risk sleep medication poses of patients developing serious side effects (i.e., dizziness, drowsiness, addiction, and relapse when medication is discontinued). Also, there is insufficient evidence for long-term effects of pharmacotherapy, and therefore long-term use of pharmacotherapy is not recommended [11–14]. Based on this lack of evidence for pharmacotherapy's long-term benefits, CBT-I is now regarded the better choice in the long-term [13]. In other words, it is assumed that CBT-I has long-term effects. Several literature reviews indeed report CBT-I’s effects may be durable [10,15,16], but they lack a quantitative data synthesis of recent evidence. To our knowledge, no meta-analysis has been published.
that included controlled studies reporting on long-term effects of (partial) CBT-I. Thus, it is unclear what the actual long-term effects of CBT-I are.

In the present meta-analysis, we aim to fill this gap in the literature by including all available RCTs reporting on the controlled long-term effects of CBT-I (at three, six and 12 mo) and quantifying these long-term effects of CBT-I. The focus of the present meta-analysis is on subjective sleep outcomes, in terms of both self-reported symptoms and sleep diaries.

**Method**

**Protocol**

Details of the protocol for this meta-analysis were registered on PROSPERO and can be accessed at www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42018094459.

**Search strategy**

Using our previous meta-analysis [9] as a starting point, we first checked whether the included studies (n = 87) had reported controlled follow-up measurements in the original papers or had published follow-up data since. This process included screening the original papers for mention of follow-up measurements done or planned, and literature search for publications from the different research groups in the five years following the first publication, using search terms “long,” “long-term” and “follow-up”. Additionally, we performed searches in Web of Science, Pubmed and PsycINFO for publications by the authors of the original article on short-term effects to see if since then long-term effects had been published.

We then performed a new search, covering the period from the end of the search (December 2015) of the previous meta-analysis [9] until May 2018. An extensive literature search was carried out in PubMed, PsycINFO, EMBASE and the Cochrane central register of controlled trials, using the same search strategy as in the previous meta-analysis [9]. Terms indicative of insomnia (i.e., insomnia, sleep disorders, sleep initiation and maintenance disorders) were paired with terms indicating psychological treatments (i.e., psychotherapy, cognitive therapy, behavior therapy). For example, our PsycINFO database query was “(DE=(“sleep disorders” or “insomnia”)) and(DE=(“psychotherapy” or “behavior therapy”))”, and specifying we looked for results from December 2015 onwards.

Titles and abstracts were screened by two persons individually (TvdZ and LB) and then crosschecked. Records definitely not meeting criteria (e.g., not a randomized trial), not aimed at insomnia, not psychological but a biological or medical treatment, not an original research report (e.g., a meta-analysis) were excluded based on title and/or abstract. We retrieved the full papers of the remaining 70 references. Two researchers assessed the papers independently (TvdZ and LB). When there was disagreement, the paper was discussed (TvdZ, LB, JL, AvS) until consensus was reached.

**Inclusion and exclusion criteria**

Inclusion criteria were: 1) randomized controlled trial (RCT); 2) investigating CBT-I or at least one component of it (see below); 3) in adults (18 y and older); 4) with self-reported and/or formally diagnosed insomnia complaints (see Table 1); 5) compared to a non-active control group (e.g., waitlist control, care-as-usual, or a minimal intervention (e.g., education about sleep or sleep hygiene information); 6) including data for sleep diary outcomes; 7) reporting controlled follow-up data for 12 or more weeks after post-test; 8) providing suitable data to calculate effect sizes.

The following components were identified as being part of CBT-I: relaxation (RE), sleep restriction therapy (SRT), stimulus control therapy (SC), paradoxical intention (PI), and identifying and challenging dysfunctional thoughts (about sleep), i.e., cognitive therapy (CT). All monotherapy (only one CBT-I-component, e.g., RE or SRT only) studies were included in line with the previous M-A [9], to ensure not overestimating effects by only taking full (current) CBT-I into account. All other therapies were excluded (e.g., interpersonal therapy, bright light therapy, exercise, tai chi, biofeedback). Studies on treatment in children or adolescents, on tapering medication use or aimed at treating a different mental health disorder and reporting on insomnia secondarily were also excluded.

**Data extraction**

We coded the following characteristics of the studies: 1) publication year, 2) recruitment setting (community, primary care, other care facilities, university), 3) the insomnia definition used, 4) comorbidity (e.g., insomnia in breast cancer patients), 5) age group...
Table 1
Characteristics of the included studies on (elements of) CBT for insomnia.

<table>
<thead>
<tr>
<th>Study</th>
<th>Recruitment</th>
<th>Definition insomnia</th>
<th>FU (3, 6 and/or 12 mo)</th>
<th>Total sample size at baseline⁴</th>
<th>Completed follow-up⁴</th>
<th>Co-morbidity</th>
<th>Age</th>
<th>Intervention</th>
<th>format</th>
<th>F2F sess</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alessi 2016</td>
<td>Comm (veterans)</td>
<td>ICSD-2 criteria</td>
<td>6 m + 12 m</td>
<td>159</td>
<td></td>
<td>FU1: Tx 92/106 (87%); C 52/53 (98%); FU2: Tx 89/106 (84%); C 51/53 (96%)</td>
<td>OSA or severe mental disorder excluded</td>
<td>60+</td>
<td>CBT-I</td>
<td>Indiv Group</td>
<td>2 F2F, 2 phone</td>
</tr>
<tr>
<td>Arnedt 2013</td>
<td>Care</td>
<td>wake time &gt; 60 m + SE &lt; 85%</td>
<td>3 m</td>
<td>33</td>
<td>Tx 15/15 (100%); C 12/15 (80%)</td>
<td>Excluded</td>
<td>18–65</td>
<td>CBT-I</td>
<td>Phone</td>
<td>4–8</td>
<td>Info</td>
</tr>
<tr>
<td>Borkovec &amp; Weerts 1976</td>
<td>Univ</td>
<td>Average SOL ≥ 30 min</td>
<td>12 m</td>
<td>24</td>
<td>n/a (Tx: n = 11; C: n = 5)</td>
<td>Allowed</td>
<td>–</td>
<td>Relaxation</td>
<td>Group</td>
<td>4</td>
<td>No txt</td>
</tr>
<tr>
<td>Casault 2015</td>
<td>Care</td>
<td>ISI ≥ 8</td>
<td>3 m + 6 m</td>
<td>38</td>
<td>FU1 and FU2: Tx: 15/20 exp (75%); C: 16/18 (88.89%)</td>
<td>Cancer patients</td>
<td>18–75</td>
<td>CBT-I</td>
<td>Self-help + phone consult</td>
<td>n/a No txt</td>
<td></td>
</tr>
<tr>
<td>Creti 2005</td>
<td>Comm</td>
<td>No formal criteria. Poor sleepers were defined as having trouble initiating or maintaining sleep. Participants were reported to have sleep onset insomnia (&gt;30 min of undesired wakefulness &gt;2 times per week, problem duration &gt;6 mo; 10% of sample), sleep maintenance insomnia (duration of awakenings after sleep onset &gt;30, &gt;2 times per week, problem duration &gt; 6 mo; 49% of sample), or both (41% of sample).”</td>
<td>12 m</td>
<td>27</td>
<td>51% of the sample</td>
<td>Excluded</td>
<td>55+</td>
<td>Relaxation</td>
<td>Self-help</td>
<td>n/a No txt</td>
<td></td>
</tr>
<tr>
<td>Currie 2000</td>
<td>Care</td>
<td>DSM-III insomnia diagnosis with SIS-D and ICSD diagnosis</td>
<td>3m</td>
<td>60</td>
<td>Tx: 28/32 (88%); C 23/28 (82%)</td>
<td>Pain</td>
<td>&lt;60</td>
<td>CBT-I</td>
<td>Group</td>
<td>7</td>
<td>WL</td>
</tr>
<tr>
<td>Edinger 2005</td>
<td>Comm</td>
<td>mean WASO ≥ 60 m</td>
<td>6 m</td>
<td>36</td>
<td>Tx: 6/18 (33%); C-SH: 7/18 (39%); C-TAU 7/11 (64%)</td>
<td>Fibromyalgia</td>
<td>21–65</td>
<td>Behavioral</td>
<td>Indiv</td>
<td>6</td>
<td>SH/TAU</td>
</tr>
<tr>
<td>Edinger 2009</td>
<td>Care</td>
<td>mean SOL + WASO ≥ 60 m</td>
<td>6 m</td>
<td>81</td>
<td>Tx: 33/41 (80%); C: 33/40 (83%)</td>
<td>Excl (PI)</td>
<td>–</td>
<td>Behavioral</td>
<td>Indiv</td>
<td>4</td>
<td>Info</td>
</tr>
<tr>
<td>Espie 2001</td>
<td>Care</td>
<td>ICSD difficulty falling/maintaining sleep, ≥ 4 n/w, ≥ 3 mo + PSQ ≥ 5</td>
<td>12 m</td>
<td>139</td>
<td>Tx: 78% of the total sample</td>
<td>Excl (CMI)</td>
<td>–</td>
<td>CBT-I</td>
<td>Group</td>
<td>6</td>
<td>WL</td>
</tr>
<tr>
<td>Espie 2007</td>
<td>Care</td>
<td>ICSD/DSM-IV criteria of insomnia</td>
<td>6 m</td>
<td>201</td>
<td>Tx: 76/107 (71%); C 67/94 (71%)</td>
<td>Allowed</td>
<td>–</td>
<td>CBT-I</td>
<td>Group</td>
<td>5</td>
<td>No txt</td>
</tr>
<tr>
<td>Espie 2008</td>
<td>Care</td>
<td>SOL or WASO ≥ 30 m, ≥ 3 n/w, ≥ 3 mo + PSQ ≥ 5</td>
<td>6 m</td>
<td>150</td>
<td>Tx: 67/100 (67%); C: 39/50 (78%)</td>
<td>Cancer</td>
<td>18+</td>
<td>CBT-I</td>
<td>Group</td>
<td>5</td>
<td>No txt</td>
</tr>
<tr>
<td>Friedman 2000</td>
<td>Comm</td>
<td>SE &lt; 80%, SOL ≥ 30m, TST &lt; 6hr, WASO ≥ 30 m, ≥ 5 n/2w</td>
<td>3 m</td>
<td>39</td>
<td>Tx: 88%; C: 100%</td>
<td>Excluded</td>
<td>55+</td>
<td>Behavioral</td>
<td>Indiv</td>
<td>5</td>
<td>Info</td>
</tr>
<tr>
<td>Fuller 2016</td>
<td>Pharmacy care</td>
<td>ISI ≥ 8</td>
<td>3 m</td>
<td>50</td>
<td>Tx: 19/22 (86%); C: 17/28 (61%)</td>
<td>Terminal illness excluded</td>
<td>18+</td>
<td>Behavioral</td>
<td>Indiv</td>
<td>4 weekly session (2 F2F, 2 phone)</td>
<td>TAU</td>
</tr>
</tbody>
</table>

(continued on next page)
<table>
<thead>
<tr>
<th>Study</th>
<th>Recruitment</th>
<th>Definition insomnia</th>
<th>FU (3, 6 and/or 12 mo)</th>
<th>Total sample size at baseline</th>
<th>Completed follow-up</th>
<th>Co-morbidity</th>
<th>Age</th>
<th>Intervention</th>
<th>format</th>
<th>F2F sess</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irwin 2014 [46]</td>
<td>Comm</td>
<td>SOL or WASO ≥ 3 n/w, ≥ 3 mo + daytime imp</td>
<td>6 m + 12 m</td>
<td>75</td>
<td>Tx: 46/50 (92%); C: 23/25 (92%)</td>
<td>Excluded</td>
<td>55+</td>
<td>PE + SC + CT + Relaxation</td>
<td>Group</td>
<td>16</td>
<td>Info</td>
</tr>
<tr>
<td>Jansson 2012 [42]</td>
<td>Care</td>
<td>SOL or WASO &gt; 30 m, ≥ 3 n/w, ≥ 6 mo + daytime imp</td>
<td>3m</td>
<td>32</td>
<td>Tx: 15/17 (88%); C: 15/15 (100%)</td>
<td>Hearing impairment</td>
<td>18–65</td>
<td>CBT-I</td>
<td>Indiv</td>
<td>7</td>
<td>WL</td>
</tr>
<tr>
<td>Jernelov 2012 [43]</td>
<td>Comm</td>
<td>ISI &gt; 10 + poor sleep ≥ 4 wk</td>
<td>3 m</td>
<td>133</td>
<td>Tx1: 41/44 (93%); Tx2: 39/45 (87%); C: 39/44 (89%)</td>
<td>Excluded</td>
<td>18+</td>
<td>CBT-I</td>
<td>Self-help</td>
<td>N/A</td>
<td>WL</td>
</tr>
<tr>
<td>Jungquist 2010 [22]</td>
<td>Care</td>
<td>SOL or WASO &gt; 30 m, &gt; 3 n/w, &gt; 6 mo</td>
<td>3 m + 6 m</td>
<td>28</td>
<td>Both FU's: Tx: 15/19 (79%); C: 5/9 (56%)</td>
<td>Pain</td>
<td>25+</td>
<td>CBT-I</td>
<td>Indiv</td>
<td>8</td>
<td>Plac</td>
</tr>
<tr>
<td>Kaldo 2015 [49]</td>
<td>Comm</td>
<td>Difficulty initiating or maintaining sleep + daytime imp + ISI &gt; 10</td>
<td>6 m + 12 m</td>
<td>148</td>
<td>Tx: 54/73 (74%); C: 53/75 (71%)</td>
<td>Excluded</td>
<td>18+</td>
<td>CBT-I</td>
<td>Self-help</td>
<td>n/a</td>
<td>Plac</td>
</tr>
<tr>
<td>Lacks 1983 [29]</td>
<td>Comm</td>
<td>WASO ≥ 30 m, ≥ 1 n/w, ≥ 6 mo</td>
<td>3 m</td>
<td>64</td>
<td>Tx: 7/15 (47%); C: 8/16 (50%)</td>
<td>Excluded</td>
<td>17–59</td>
<td>SC</td>
<td>Group</td>
<td>4</td>
<td>Plac</td>
</tr>
<tr>
<td>Lichtstein 2000 [32]</td>
<td>Comm</td>
<td>SOL or WASO ≥ 30 m, ≥ 3 n/w, ≥ 6 mo + daytime imp</td>
<td>3 m</td>
<td>44</td>
<td>Tx: 22/23 (96%); C: 17/21 (81%)</td>
<td>Illness</td>
<td>58+</td>
<td>Relaxation + SC</td>
<td>Indiv</td>
<td>4</td>
<td>WL</td>
</tr>
<tr>
<td>Lichtstein 2001 [33]</td>
<td>Comm</td>
<td>SOL or WASO ≥ 30 m, ≥ 3 n/w, ≥ 6 mo + daytime imp</td>
<td>12 m</td>
<td>89</td>
<td>83% of the total sample</td>
<td>Excluded</td>
<td>59+</td>
<td>Relaxation + SR</td>
<td>Indiv</td>
<td>6</td>
<td>Plac</td>
</tr>
<tr>
<td>Lovato 2014 [47]</td>
<td>Comm</td>
<td>WASO &gt; 30 m, ≥ 3 n/w, &gt; 6 mo + daytime imp</td>
<td>3 m</td>
<td>118</td>
<td>Tx: 72/86 (84%); C: 27/32 (84%)</td>
<td>Excluded</td>
<td>Older</td>
<td>CBT-I</td>
<td>Group</td>
<td>4</td>
<td>WL</td>
</tr>
<tr>
<td>McCurry 2016 [54]</td>
<td>Comm</td>
<td>ISI &gt; 10 + poor sleep ≥ 4 wk</td>
<td>3 m</td>
<td>106</td>
<td>Tx: 44/53 (83%); C: 42/53 (79%)</td>
<td>Menopausal</td>
<td>40–65</td>
<td>CBT-I</td>
<td>Phone</td>
<td>n/a</td>
<td>Menopause education</td>
</tr>
<tr>
<td>Morin 2005 [37]</td>
<td>Comm</td>
<td>Diagnostic criteria for insomnia: (1) complaint of poor sleep quality or dissatisfaction regarding sleep; (2) symptoms of initial, maintenance, or late insomnia at least 3 nights per week; (3) presence of psychological distress or daytime impairment related to sleep difficulties; and (4) presence of the sleep difficulties for at least 1 mo.</td>
<td>6 m</td>
<td>192</td>
<td>Tx: 81/96 (84%); C: 86/96 (90%)</td>
<td>Allowed</td>
<td>18+</td>
<td>CBT-I</td>
<td>Self-help</td>
<td>n/a</td>
<td>No txt</td>
</tr>
<tr>
<td>Rybarczyk 2002 [34]</td>
<td>Care</td>
<td>SOL ≥ 45 m or WASO ≥ 60 m or TST ≤ 5 h, ≥ 3 n/w</td>
<td>3 m</td>
<td>24</td>
<td>Tx: 10/16 (63%); Tx2: 13/18 (72%); C: 12/17 (71%)</td>
<td>Illnesses</td>
<td>55+</td>
<td>CBT-I</td>
<td>Group/ Self-help</td>
<td>8 n/a</td>
<td>WL</td>
</tr>
<tr>
<td>Savard 2016 [48]</td>
<td>Care</td>
<td>ISI ≥ 8 or ≥ 2 nights of sleep medication in last 2 wk</td>
<td>3m + 6m + 12m</td>
<td>242</td>
<td>Tx1: 61/81 (75%); Tx2: 49/80 (61%); C: 49/81 (60%)</td>
<td>Cancer</td>
<td>18–75</td>
<td>CBT-I</td>
<td>Indiv/video</td>
<td>6 n/a</td>
<td>No txt</td>
</tr>
</tbody>
</table>
Quality assessment

Using the criteria suggested in the Cochrane handbook [17], we assessed the validity of the studies: 1) adequate sequence generation, 2) concealment of allocation, 3) adequate handling of incomplete outcome data, and 4) selective reporting of data. We did not assess the binding of patients or therapists since this is not possible in psychotherapy research nor did we assess binding of outcome assessors since all reported outcomes are based on self-report. Two reviewers conducted the quality assessment independently of each other (TvdZ and LB) and then crosschecked their findings.

Meta-analysis

We focused on the effects on insomnia severity primarily (measured through questionnaires), and secondarily on sleep onset latency (SOL) and sleep efficiency (SE) measured through sleep diaries because these are among the most important sleep outcomes and yielded the largest number of comparisons. We defined an effect as “long-term” if it was measured at least 12 wk after the end of treatment. We created three subgroups, grouped in 1) three months (between 12- and 17.5-wk post-treatment), 2) six months (between 18 and 35 wk) and 3) 12 mo (more than 35 wk) for pragmatic reasons. There were no studies reporting outcomes after one year. Where the exact timing of the follow-up assessment was unclear from the articles (i.e., post-baseline or post-treatment), we assumed the reported follow-up period was post-treatment (however, this assumption did not influence the group assignment of this study).

We computed Hedges’ g to determine between group effect sizes. Hedges’ g is a measure of standardized mean differences (similar to Cohen’s d), after adjusting for small sample sizes [18,19]. This effect size (based on the differences between conditions at the different time points) can be interpreted as the difference between the mean scores of the two groups expressed in the number of weighted pooled standard deviations. Effect sizes (ES) are commonly interpreted as either large (>0.56), moderate (0.33–0.55) or small (0–0.32) [20]. The available statistics (means, standard deviations, standard errors, 95% confidence intervals, and interquartile ranges) from the included studies were used to compute the ESs using the metafor package in R [21].

We checked for outliers by visually inspecting forest plots for all analyses. We defined outliers as studies in which we found a 95% confidence interval around the ES that did not show overlap with the 95% confidence interval of the pooled effect size. One study was identified to be an outlier [22] on all variables it included (SOL, WASO, SE, TST and ISI at 3 mo and 6 mo). Based on author consensus we removed this study from the analysis because we judged it to be a significant and pronounced outlier.

In the 12 mo follow-up, one study had to be excluded from the analysis because means were not provided and could not be
calculated from what was reported (WASO and TST) [23]. For SOL, two studies did not report a measure of dispersion [23,24], but effect sizes were calculated using other available statistics.

Expecting the studies to show significant heterogeneity, we used the random effects model in the metafor package [21] to estimate the weighted pooled effect sizes for the different outcome measures. Heterogeneity was inspected using the fixed effects model using I², describing the variance between studies as a proportion of the total variance. A value of 0% means there is no observed heterogeneity, with larger percentages indicating more heterogeneity [25]. The 95% confidence intervals around I² were calculated.

To address treatment heterogeneity in the included studies we performed sensitivity analyses, exploring differences in effects of full CBT-I compared to effects of partial CBT-I or monotherapy.

To assess potential publication bias, we visually inspected the funnel plot and conducted Egger’s test for all variables, at the three different time points [26]. The Duval and Tweedie [27] trim and fill procedure was used to adjust the effect size for publication bias and to provide an indication of the number of studies that might have been missing from the analysis.

Results are reported in accordance with the PRISMA statement and the PRISMA checklist can be found in the supplement [28]. Results in the expected and desired direction are reported as positive results, results in the undesired direction (i.e., worsening of symptoms) are reported as negative results.

Results

Selection of included studies

We selected studies using two strategies. First, we included articles identified in the previous meta-analysis [9], which had controlled 12 wk or more follow-up data. A total of 25 studies were found using this method [23,24,29–50]. Secondly, we performed a new literature search for the period since the previous meta-analysis [9]. We screened 420 titles and abstracts and excluded 324 as not relevant and/or not meeting criteria based on their titles and abstracts. In total, we then assessed 96 full texts. Of these, we excluded 66 articles because 1) duplicates between the first and second search strategy (n = 5), 2) no RCT (n = 24), 3) no CBT component (n = 3), 4) no sleep diary (n = 16), 5) no treatment (n = 14), 6) no follow-up (n = 3), 7) no data we could use for calculations of the effect sizes (n = 5) or 8) not a results paper (n = 3). The remaining 30 papers were included in this meta-analysis: the 25 already included in our previous meta-analysis and five new papers reporting follow-up results of an RCT on (a component of) CBT-I compared to an inactive control group [51–55]. A flowchart of the inclusion and exclusion process is in Fig. 1.

Two studies compared different active interventions in comparison to a control condition. Therefore, the meta-analysis included 32 comparisons in total. Some studies (n = 9) reported multiple follow-up measurements and not all studies reported the same sleep variables. The numbers of comparisons vary in the different analyses and are listed in the tables.

Study characteristics

Eighteen studies reported three months follow-up data [22,29–32,34,38,42–45,47,48,50,51,53–55], fourteen studies reported data for six months follow-up [22,36–41,46,48–52,55] and eight studies reported 12 mo follow-up data [23,24,33,35,46,48,49,52]. The oldest study in our sample was published in 1976 [24]. The large majority was published after 2000 (n = 28; 93%). Fourteen studies (47%) investigated community samples, 15 studies (50%) recruited patients from care settings and one study recruited within a university (3%). Around 50% of the studies (n = 16) excluded patients showing (specific) comorbidities. The remaining 14 studies did not, or did not report it. Out of the 30 studies included in the qualitative assessments, twenty studies (67%) offered full CBT-I, the others (n = 10; 33%) offered one or more components of CBT-I but not the full package. Treatment was offered in group format in 13 studies, in individual format in 13 studies and in self-help format in six studies (total n = 32 treatment groups across studies). Treatment groups were compared to waitlist (n = 8), no treatment (n = 7), care-as-usual (n = 2), placebo (n = 6) or minimal intervention (n = 8) control groups (total n = 31 control groups across studies). See Table 1 for details.

Quality assessments

Sixteen out of the 30 studies included in the qualitative assessments (53%) generated randomization sequences adequately, whereas 14 (47%) did not, or did not report on it. A total of 13 studies (43%) reported adequate concealing of the random allocation, for the other 17 studies (57%) this was not reported. Twenty studies (67%) reported handling missing data by performing intent-to-treat analysis, six studies (20%) did not conduct intent-to-treat analysis and four (13%) did not report how they handled missing data. Eight studies (27%) seemed to be selective in reporting of data based on comparisons of study protocols to results reported.

Three to 12 mo effects on insomnia severity, sleep efficiency and sleep onset latency

Three months after CBT-I finished the effect compared to the non-treated controls on ISI was statistically significant and large (g = 0.64; Nc = 13). The effect remained present after three months, with a moderate between group effect size at six months (g = 0.40; Nc = 8) and a small between group effect size at 12 mo (g = 0.25; Nc = 4).

Three months after CBT-I was finished the effect on sleep efficiency was moderate (g = 0.51; Nc = 21) and remained relatively stable over time (six months: g = 0.32; Nc = 16; 12 mo g = 0.35; Nc = 8). The effect size of CBT-I on SOL declined from moderate at three months to small at six months but returned to moderate at 12 mo (g = 0.38, 0.29 and 0.40 respectively with Nc = 21, 16, and 10). Forest plots for ISI, SE and SOL can be found in Figs. S1–S3.

Three to 12 mo effects on secondary outcomes

For the secondary variables, treated groups outperformed the control groups at three months on most variables (PSQI: g = 0.80; Nc = 6; WASO: g = 0.42; Nc = 20; SQ g = 0.49; Nc = 5). Effects on TST (g = 0.06; Nc = 21) and NWAK (g = 0.08; Nc = 4) were not significant (see Table 2). At six months, effects were present, but most had become slightly smaller (WASO: g = 0.27; Nc = 13; PSQI: g = 0.48; Nc = 3). Effects on SQ (g = 0.09; Nc = 2) and TST (g = 0.05; Nc = 15) were non-significant. NWAK data were not available at six months. At 12 mo, results are less reliable due to the small number of studies included in the analysis (WASO: g = 0.26; Nc = 8; NWAK: g = 0.52; Nc = 2). Effects on PSQI, SQ and TST were not significant (PSQI: g = 0.22; Nc = 2; SQ: g = 0.24; Nc = 3; TST: g = 0.03; Nc = 7).

Differences between post-test effects of studies with long-term effects vs all studies

We calculated the post-test effect sizes of the 29 studies (excluding one outlier) in the current meta-analysis (ISI, SE, SOL).
We then divided the studies based on follow-up length and compared the post-test effects of studies with 3 mo follow-up, 6 mo follow-up and 12 mo follow-up to the overall post-test effect sizes (results are in Supplementary Table S1).

Sensitivity analysis

We made a comparison between the sample with all studies included to the sample with only the studies with full CBT-I. Due to variety in follow-up length and reported variables there were different samples only for ISI (3m), SOL (3/6/12m), and SE (3/12m; see Supplemental Table S2 for Nc’s and effect sizes). The differences in \( g \) ranged from 0.01 to 0.05. We also compared the sample with all studies included with the sample excluding monotherapies. There were different samples only for SE (Nc = 8 versus 6) and SOL (Nc = 10 vs 8), both at 12 m. The difference in \( g \) was 0.03 for both (see Supplemental Table S2).

Publication bias

We found indications for publication bias in only one of the analyses performed: the three-month follow-up effects on total sleep time. Egger’s test for funnel plot asymmetry was significant \( (t = 2.38, df = 19, p = .028) \). We performed a trim and fill analysis and found the estimated number of studies missing on the left side of the funnel plot to be \( n = 1 \). The effect size adjusted for publication bias for three-month TST was 0.039 \( (p = .601) \), slightly smaller than the 0.061 \( (p = .391) \) unadjusted effect size, and remained statistically non-significant.

Discussion

We performed a meta-analysis on 29 RCTs (removing one study which was an outlier) to investigate the long-term effects of cognitive and/or behavioral treatments for insomnia. We studied three, six and 12-mo follow-up data in separate analyses. Three months after treatment, the severity of insomnia complaints (primary outcome) was considerably better for patients treated with CBT-I than for patients without an active treatment (ISI: \( g = 0.64 \)). Effects were of moderate size at six months (ISI: \( g = 0.40 \)). At 12 mo the treated group still outperformed the control group, albeit showing a smaller effect size (ISI: \( g = 0.25 \)). The sleep diary variables (sleep efficiency and sleep onset latency; secondary outcomes) showed a similar pattern, although effect sizes were smaller.
Table 2
Effects of insomnia treatments at 3-, 6- and 12-mo follow-up (outlier removed).

<table>
<thead>
<tr>
<th>Metric</th>
<th>3 mo follow-up</th>
<th>6 mo follow-up</th>
<th>12 mo follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nc</td>
<td>95% CI</td>
<td>95% CI</td>
<td>95% CI</td>
</tr>
<tr>
<td>I²</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 wk post-treatment</td>
<td>0.86 (0.56)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td>18 wk post-treatment</td>
<td>0.56 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td>35 wk post-treatment</td>
<td>0.38 (0.20)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td>40 wk post-treatment</td>
<td>0.35 (0.21)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
</tr>
</tbody>
</table>

Note. g = Hedge’s g; SE = sleep efficiency; WASO = wake after sleep onset; SOL = sleep onset latency; SQ = sleep quality; NWAK = number of awakenings; TST = total sleep time; Nc = number of comparisons.

Adjustment is for multiple comparisons using Bonferroni correction. Differences between pre- and post-treatment were assessed using paired t-test. Differences between 12 wk and 18 wk were assessed using unpaired t-test.

For example, the symptom decrease in control participants is unlikely to have happened spontaneously. A three-year follow-up study by Blom and colleagues showed that during the follow-up phase, patients have sought effective help elsewhere, this may have led to a smaller difference between the treatment conditions and the controls.

A second possible reason is that the patients in the control conditions start sleeping better over time, either because they have sought treatment elsewhere or merely as a result of time passing. When looking at the data from the individual studies included in this meta-analysis, we do observe control participants reporting a small decrease in insomnia symptoms. We do not know whether control patients have undergone treatment. As the time of follow-up increases, so does the window of opportunity to seek treatment elsewhere. Earlier research on the natural course of insomnia reports that insomnia is often persistent when untreated [56], suggesting the symptom decrease in control participants is unlikely to have happened spontaneously. A three-year follow-up study by Blom and colleagues showed that during the follow-up phase, control group participants were indeed likely to use sleep medication and seek additional insomnia treatment elsewhere [57]. Our results may therefore be on the conservative side: if control participants have sought effective help elsewhere, this may have led to a smaller difference between the treatment conditions and the controls.

A third possible reason is an increase in the intervention participants’ insomnia symptoms over time. Again, when looking at the data for the individual studies included in this meta-analysis, we see that on average intervention participants report a slight return of symptoms over time. This would mean the effects of the intervention in the long run are not as pronounced as is often thought. This explanation seems plausible: the interventions are relatively short, and focus on behavioral changes (e.g., lifestyle, bedtimes, sleep hygiene). These behavior changes can be hard to stick to, as we know from extensive research in other areas of preventive care where lifestyle changes are necessary (e.g., increasing physical activity, smoking cessation, reducing alcohol consumption [58]).
Our results, therefore, may be explained by an interplay between these three explanations. To answer this question more definitively, we need more randomized controlled trials that include a long and controlled follow-up with multiple measurements (e.g., 3 mo and 12 mo) reporting on what happens in the follow-up phase as well. It would also be good to adopt a more uniform research approach in reporting on (long-term) treatment trials. For this meta-analysis, included studies showed substantial heterogeneity. Furthermore, not all variables investigated were reported in all included studies, limiting the number of comparisons. Pragmatic choices had to be made in choosing outcome variables that present a balanced picture, but SE in particular might not be an ideal outcome to report [59]. Given the small number of studies reporting 12-mo follow-up data (n = 8), results at 12 mo should be interpreted with caution. To enhance uniformity, we advise CBT-I researchers to include the Carney consensus diary variables [60] and the ISI [61] in future research. This would increase the opportunities for meaningful pooling of evidence and also allow meaningful subgroup analyses providing insight into the variables influencing magnitude of effect size.

The heterogeneity of included studies is a potential limitation of this meta-analysis. Studies differ in terms of comorbid populations, delivery mode and treatment content. We decided to include these studies for several reasons: 1) Current diagnostic practice no longer makes a distinction between primary and secondary/comorbid insomnia. 2) Previous research shows delivery mode is not an important factor determining effects [62,63]. 3) The meta-analysis by van Straten and colleagues (2018) showed full CBT-I effects did not differ from those of partial CBT-I [9]. Sensitivity analysis on the current dataset supported our choice (Supplemental Table S2).

Importantly, although the effects decline somewhat over time, we did find sustained long-term CBT-I effects. These established long-term effects strengthen the claim that CBT-I outperforms pharmacotherapy in the long run and is the preferred treatment for insomnia (e.g., [13]). The sustained and (tentatively) clinically relevant effects (ES > 0.25 for depression treatment according to Cuijpers and colleagues [64]) are of particular interest. It must be noted however that relatively few meta-analyses on other psychological disorders have established long-term controlled treatment effects [65–67]. Meta-analyses are often limited to short-term effects, due to difficulties interpreting varied follow-up intervals and potential for other treatments or life events during the longer follow-up phase [67]. The few meta-analyses that have investigated long-term effects have generally reported similar findings: sustained (but somewhat declined) long-term treatment effects for CBT for depression [68,69], anxiety [70] and PTSD [71].

To enhance long-term effects, perhaps we need to put more emphasis on relapse prevention within CBT-I. Currently, relapse prevention is a component of most interventions at the end of treatment, making patients aware of the potential of relapse. They are advised to return to the exercises in the intervention when this happens. This could be improved, for example, by adding a booster session after six months or asking patients to continue keeping a sleep diary for a longer period of time.

Taken together, the results of the present meta-analysis show favorable effects of CBT-I at follow-up, still present (albeit smaller) at 12 mo after treatment. Establishing this long-term effect is of major importance: it provides a strong argument for the clinical recommendation of offering CBT-I. Ultimately, the main goal of CBT-I is to improve daytime functioning, improving quality of life and reducing societal costs due to absenteeism and work productivity losses. Research on long-term effects of CBT-I on these measures is currently scarce, although daytime impairment is the main reason patients seek treatment [2,72]. This meta-analysis indicates that CBT-I does show the often-claimed long-term effectiveness, but it is not without its limitations. In further insomnia research, we need to aim for more uniformity and more controlled studies with a longer follow-up (preferably one year). This would enable confidently stating that CBT-I has the often-proclaimed long term effects, improves both insomnia severity and daytime functioning and should be the first treatment of choice, in line with recent recommendations for the treatment of insomnia symptoms [12,13].

Author contributions
TZ, AVS, SD and JL conceived of and designed the study. TZ, AVS and JL performed the literature search. TZ and LB screened the search results, completed the inclusion and exclusion process, and extracted and cross-validated the data, supervised by AVS and JL. TZ and LB drew up the analysis plan under supervision of AVS and JL. LB coded the R-scripts under supervision of TZ. TZ drafted the manuscript. All authors contributed to and approved the manuscript.

Practice points
- Cognitive behavioral therapy for insomnia is effective on insomnia severity, sleep efficiency and sleep onset latency at three, six- and 12-mo follow-up.
- Long-term effects provide support for international guidelines recommending cognitive behavioral therapy as the first treatment option for insomnia.

Research agenda
Future studies on cognitive behavioral treatment for insomnia should aim to:
- Include long-term controlled follow-ups
- Provide information on daytime functioning as a consequence of insomnia and changes in functioning after treatment
- Be uniform in reporting on randomized controlled trials and adequately monitor control participants during the follow-up phase.

Acknowledgements
All authors declare that they have no competing interests and did not receive financial support to perform this meta-analysis. SDK is supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC).

Appendix A. Supplementary data
Supplementary data to this article can be found online at https://doi.org/10.1016/j.smrv.2019.08.002.
References


[31] * The most important references are denoted by an asterisk.