Comparative Effectiveness Review Disposition of Comments Report

Research Review Title: *Management of Insomnia Disorder*

Draft review available for public comment from November, 2014 to December 2014.

Research Review Citation: Brasure M, MacDonald R, Fuchs E, Olson CM, Carlyle M, Diem S, Koffel E, Khawaja IS, Ouellette J, Butler M, Kane RL, Wilt TJ. Management of Insomnia Disorder. Comparative Effectiveness Review No. 159. (Prepared by the Minnesota Evidence-based Practice Center under Contract No. 290-2012-00016-I). AHRQ Publication No.15(16)-EHC027-EF. Rockville, MD: Agency for Healthcare Research and Quality. December 2015. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail or E-mail. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the EHC Program Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2163
Published Online: December 30, 2015

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Introduction	The PSQI is not an appropriate outcome measure or measure of insomnia treatment effects on function.	Our systematic review was restricted to the instruments used in the original research and the PSQI was relatively frequently used.
Peer Reviewer 1	Introduction	Withdrawal is not at all a specific and adequate measure of harms, which should be examined in detail.	Our systematic review was restricted to harms reported in the original research. Many studies did not report specific harms and withdrawals were the best proxy available to us. Specific harms were reported inconsistently; we extracted withdrawals, withdrawals to adverse effects, and number of subjects reporting more than one harm when these data were reported to capture an aggregate assessment of these harms enabling us to assess harms across the body of evidence.
TEP Reviewer 2	Introduction	No specific comments.	
TEP Reviewer 3	Introduction	Page 2 Paragraph 2 'AASM' Has not been defined prior to the abbreviation	Thank you, we've corrected this oversight: 'The American Academy of Sleep Medicine (AASM)'.
TEP Reviewer 3	Introduction	Page 4 Table 2 in both main document and executive summary The definition of CBT can be improved: 'Combination treatments that include cognitive and behavioral Components' can be changed to: A combination of treatments that include cognitive restructuring around sleep and behavioral modifications (sleep restriction, stimulus control) and education (sleep hygiene)	Thank you, we've replace our description with your suggestion in Table A (p. ES-2) and Table 2 (p. 4).
Peer Reviewer 4	Introduction	The requirement that all pharmaceutical interventions included are licensed for use in the USA means that papers describing medications which are used extensively elsewhere in the world are excluded. This may be because the manufacturers have not applied to the FDA or that at the time of application there was not enough evidence for their use, but subsequently there may be better evidence. Once a drug is out of patent there is no incentive to finance proper modern studies. It is unclear in the review why this requirement was adhered to – I am assuming it was in the remit that the researchers were given by the US government. It means that when this review is published (which I hope it will be, in a good sleep journal) and it is accessed all over the world, it will not necessarily be applicable worldwide and this needs to be made clear in the text.	Correct, AHRQ reviews focus on interventions currently available in the United States and our results do not include interventions available in other parts of the world (i.e. zopiclone, idiplon).

Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2163
Published Online: December 30, 2015

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 6	Introduction	The Introduction could make a few additional statements about the fact that insomnia is often undiagnosed and remains untreated. Also, while it has historically been conceptualized as a situational symptom, there is now evidence that insomnia can be a disorder of its own and it is often a persistent problem over time (Morin et al., 2009; Arch Int Med). In addition, there is increasing evidence that persistent insomnia increases risks for depression (Baglionni et al., 2012) and hypertension (Laustand et al., 2012 or 2013). These important issues warrant some discussion to highlight the importance of treating this common sleep disorder.	Added text: 'Insomnia is often not diagnosed and may remain untreated.' 'Insomnia disorder is associated with medical and psychiatric morbidity including hypertension and depression.'
TEP Reviewer 8	Introduction	This is succinct, accurate, and appropriate in scope and detail.	Thank you.
TEP Reviewer 8	Introduction	I would not characterize sleep efficiency as a more 'comprehensive' sleep measure (ES-2). In fact, it is both a derivative measure and a ratio, which could arguably lead to it being a less desirable outcome.	Removed 'a more comprehensive sleep measure'.
TEP Reviewer 8	Introduction	I'm not sure I would call daytime functional outcomes 'secondary' (ES-2).	Changed wording to exclude 'secondary': 'Insomnia treatments can also improve patient-centered outcomes such as mood and well-being, quality of life, and productivity through improved sleep.'
TEP Reviewer 8	Introduction	Table A includes cognitive behavioral therapy twice. Table A does not use the same terms as Tables 1 and 2.	Removed second 'cognitive behavioral therapy' line from Table A. Revised terminology in Tables 1 and 2 for consistency.
TEP Reviewer 9	Introduction	Intro p.3. Table 1 should include Relaxation therapy under Psychological treatments.	Added relaxation therapy under psychological treatments in Table 1.
Peer Reviewer 10	Introduction	This is a bit confusing - AASM recs are supported 'by the highest quality evidence' but this review suggests evidence is low to moderate. Please clarify whether you mean that AASM claimed that their recs were based on highest quality evidence (and you disagree) or if this actually refers to your finding that, in general, there may have been stronger evidence for CBT than other treatments.	AASM evidence quality and AHRQ strength of evidence are different. We removed the statement to avoid confusion.
Peer Reviewer 10	Introduction	Table A ES-2 - cognitive behavior rx listed twice. The definition just restates the term itself but I don't think will be informative to most non-psychologists who don't do CBT. What is a rx that has a cognitive component? Does this include the other things listed in the table like relaxation training and biofeedback?	Removed second 'cognitive behavioral therapy' line from Table A. Revised description of CBT to 'a combination of treatments that include cognitive restructuring around sleep and behavioral modifications (sleep restriction, stimulus control) and education (sleep hygiene)'.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Methods	On page 43, line 48-49, the review stated that 'we felt strongly that a minimum 4-week study duration was appropriate', even though FDA has not approved or recommended most hypnotics for 4 weeks, and most patients receiving hypnotics take them for shorter intervals. At the bottom of page 43, the review also acknowledged that 'a large proportion of drug trials' were shorter than 4 weeks, and most previous reviews included them. One may see that the duration-of-treatment criterion is idiosyncratic and contrary to the views of most experts. To 'capture' results of <4-week treatments from previous reviews is not what should have been done, since as stated (page 43, lines 55-56), this 'capture' is not by itself an adequate substitute for de novo extraction, among other reasons because it is inevitably out of date.	Our review focused on 'insomnia disorder' or what best maps to 'chronic insomnia'. This disorder is diagnosed when it occurs for a minimum of three months. Studying treatment for less than four weeks does not provide adequate information on the treatment of a chronic condition. Drugs that are FDA approved for short-term use may not be appropriate treatments for a chronic condition. Our goal was to assess treatments for insomnia disorder and was not a systematic review of hypnotic medications.
Peer Reviewer 1	Methods	Likewise, identifying eligible studies prior to 2004 from published reviews is not by itself a satisfactory method of extraction (page 44, lines 23-24 & 29).	We relied on previous systematic reviews to identify relevant trials published prior to 2004. When those trials were deemed eligible, we did extract them using the same methodology used to extract trials identified using the bibliographic database searching.
Peer Reviewer 1	Methods	The comparative effectiveness review failed to include barbiturates or diphenhydramine (which is a prescription drug approved for insomnia). Granted, diphenhydramine and similar drugs are often prescribed over-the-counter or for other indications, making their inclusion difficult. Diphenhydramine is listed in Table 1, page 38.	We included any drug available in the United States if there were eligible trials. We did not identify eligible trials assessing the efficacy of barbiturates or diphenhydramine.
Peer Reviewer 1	Methods	Likewise, the review failed to consider internet forms of CBTI.	We included any form of CBT-I; several trials included arms with web-based CBT-I.
Peer Reviewer 1	Methods	The comparative effectiveness review failed to discuss studies of bright light treatment in nursing homes, for which there are several controlled trials with insomnia end points.	The relevant population for our review was community-dwelling adults. Trials of institutionalized patients were excluded.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Methods	According to page 36, line35, insomnia should not arise exclusively in the course of another sleep-wake disorder; however, the PSQI contains items focused towards sleep apnea, restless legs, circadian rhythm phase disorders, and excessive sleepiness: PSQI is NOT (in the majority of PSQI items) a measure of symptoms or outcomes of insomnia. There is a consensus that PSQI scores correlate poorly with objective sleep latency, total sleep time, sleep efficiency, or awakenings. Indeed, PSQI correlates best with measures of depression. PSQI is NOT an appropriate or specific measure of global outcome of insomnia treatments.	The DSM-5 diagnostic criteria require that 'insomnia should not arise exclusively in the course of another sleep-wake disorder'. The PSQI was a frequently used instrument used in the literature and systematic reviews are limited to instruments used in the primary literature. Because it contained items assessing sleep as well as daytime functioning, we classified it as a 'global outcome'. Because we only included studies enrolling patients with insomnia, we do not believe that the specific items in the PSQI focused on other sleep-wake disorders have an important impact on our results.
Peer Reviewer 1	Methods	A participant's study withdrawal is not a sufficient measure of adverse events. Moreover, study withdrawal is not a specific measure of adverse events. As the report recognized, study withdrawal may reflect aspects of motivation and a participant's external circumstances as well as the seriousness of adverse effects.	We agree. Our systematic review assessed many harms outcomes. We assessed harms reported in the original research. Many studies did not report specific harms and withdrawals were the best proxy available to us. Specific harms were reported inconsistently; we extracted withdrawals, withdrawals due to adverse effects, and number of subjects reporting more than one harm when these data were reported to capture an aggregate assessment of these harms, enabling us to assess harms across the body of evidence.
Peer Reviewer 1	Methods	There is another kind of adverse evidence, evidence of post-drug-withdrawal insomnia, which is not mentioned at all, and may occur after withdrawal.	We did not include studies of hypnotic withdrawal. Those studies were beyond the scope of this review because they assessed adverse effects once treatment was stopped. Our focus was on response and adverse effects associated with active treatment.
Peer Reviewer 1	Methods	The previous reviews mentioned on page 33, lines 15-27, have demonstrated that it is perfectly feasible to analyze specific adverse effects of hypnotics and other insomnia treatments. The FDA has expressed great concern about daytime hangover (sedation) effects of hypnotics, particularly as this may be related to the high rate of automobile crashes among hypnotic recipients. Sleep restriction may also produce daytime sleepiness. This review should have expanded on the previous studies of daytime sedation.	Previous reviews included drug trials with durations less than four weeks, greatly expanding the number of trials and subjects with which to analyze specific adverse effects. We focused on adverse effects proxies that could be aggregated and for which strength of evidence could be assessed.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Methods	Since the review often found data concerning benefits inadequate, there is no reason why it should not study adverse effects, even should data concerning some adverse effects may likewise be found to be inadequate. When collected data on adverse events is inadequate, the reviewers should say so. Instead, the review presents an unbalanced focus on benefits and whitewashes risks and harms, though the proposed Key Question 1 and Key Question 2 weighed benefits and harms (risks) equally (page 40).	We studied adverse effects when reported. This included findings from RCTs used to assess benefits as well as longer term observational studies for which we did not assess 'benefits' data. We did not assess strength of evidence on this data. We assessed strength of evidence for withdrawals, withdrawals due to adverse effects, and subjects reporting more than one adverse effect in the RCT data because these measures were most consistently reported and therefore likely to produce the highest strength evidence. We described specific adverse effects in the text.
Peer Reviewer 1	Methods	On page 41, line 38-41, it appears that ONLY subjective measures of sleep were included, but no objective measures. The draft for public comment dated October 25, 2013 specifically called for polysomnographic as well as subjective data in outcome assessment, though this seems to have been lost in the protocol published by AHRQ on April 3, 2014. Subjective data may be 'patient centered,' but it is hardly science-centered. The well-known amnesic effects of hypnotics produce a situation where patient reports of benefit consistently exceed objective benefits, and placebo benefits are rated highly. Neither FDA nor commercial advertising of hypnotics is willing to ignore the objective data, and one doubts that primary practitioners will be adequately informed without objective data.	We elected to concentrate on sleep diary measures because they are patient-centered. Diagnosis of insomnia disorder relies on patient report and not sleep laboratory parameters and we therefore considered these measures to be most clinically relevant. Additionally, our review is aimed towards primary care practitioners who will not likely use polysomography for diagnosis or followup of their insomnia patients making subjective sleep outcomes most applicable to practice. Primary care providers are likely to value how their patient feels they are sleeping and how perceived sleep impacts their life on a routine basis.
Peer Reviewer 1	Methods	Incidentally, deaths are an extremely important outcome which should be reported in every controlled trial (even if there are no deaths): see Iannidis, Evans, Gotzsche, ...CONSORT Group, et al. 2004. Does the strategy of this review ignore deaths because no patient self-reported his own death? There were indeed deaths in these studies, e.g., Ancoli-Israel et al., reference 108, which are not mentioned.	We described specific adverse effects including deaths in the text. They do not appear in the tables because we did not assess strength of evidence on specific outcomes. We included patient-centered outcomes which are not exclusive to patient-reported outcomes. We have added any information on deaths and other serious adverse events from Ancoli-Israel and other trials when reported. We have added text to describe the death inadvertently omitted from the text.
TEP Reviewer 2	Methods	The distinction between global outcomes versus secondary outcomes is not clear. As it currently reads, global outcomes include PGI, ISI, PSQI but quality of life measures (e.g., SF-36, WHOQOL) are considered secondary outcomes. It would seem that QoL could also be conceptualized as	Our thoughts in classifying outcomes into three groups (global, sleep, and secondary) was global outcomes assessed both sleep and functioning/distress, sleep outcomes were only sleep parameters, and secondary outcomes were outcomes such as functioning and

Commentator & Affiliation	Section	Comment	Response
		global. Can these distinctions be clarified?	mood; secondary were considered 'secondary' to improved sleep (not direct consequences of the treatment). However, the terminology seemed confusing to several peer reviewers and we will call these secondary outcomes 'functioning, mood, and quality of life outcomes'.
TEP Reviewer 2	Methods	Clarification of how each investigator determined overall risk of bias in each domain would be helpful. Given that one of the conclusions is that the strength of evidence is low to moderate overall, greater specification as to what contributed to these ratings would be helpful.	We elaborated on the specific risk of bias assessments and included investigator guidance in the appendices. Revised text: 'Two investigators independently assessed risk of bias for eligible RCTs using an assessment tool developed for this project (Appendix B). {Viswanathan, 2012 #1422; Higgins 2011} Investigators assessed several types of bias including selection bias (method of randomization, groups similarity at baseline, allocation concealment), performance bias (blinding of provider and recipient, intervention definition – theory based, manualized, fidelity to treatment), detection bias (outcome assessors blinded, instruments validated and reliable, clinical significance of outcomes, co-interventions avoided or similar, correction for multiple comparisons, power – if pooling not possible), attrition bias (extent of attrition, reasons for incomplete data provided, incomplete data handled appropriately), reporting bias (select group of outcomes reported, select analysis conducted), and other sources of bias. Certain items (such as adequacy of intervention definition and implementation) were especially necessary to adequately capture all potential risk of bias associated with psychological interventions.'
TEP Reviewer 3	Methods	Inclusion/exclusion criteria are justifiable. Search strategies were explicitly stated and logical. Definitions and statistical methods were appropriate.	Thank you.
Peer Reviewer #4	Methods	Inclusion criteria OK but see above, methods v good	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 6	Methods	The selection criteria could be made more explicit. Under inclusion criteria in Table 3, we see 'exclusive subgroups of adults' (older, comorbid....). This is not clear to me whether older adutls were retained or excluded and whether all studies with medical or psychiatric comorbidities were excluded. Nothing is said about concurrent use of hypnotic medications, a common occurrence for studies of psychological interventions.	We revised Table 3 to clarify inclusion criteria. We did not explicitly exclude studies with patients concurrently taking hypnotic medications, but this would have been assessed during risk of bias assessment of the individual study.
TEP Reviewer 6	Methods	It appears that studies using only polysomnography to evaluate outcome were not retained for this review. This decision should be justified as it may have excluded many drug trials that used only PSG to evaluate outcome.	We revised text to address this comment: 'We included studies that report subjective outcomes. Polysomography outcomes are not patient-centered and trials reporting only these outcomes were excluded. Providers use history and patient report to diagnose insomnia disorder and assess patient opinion regarding treatment. Providers are more likely to value a patients' perspective of improvement based upon their typical sleep routine. Sleep parameters obtained in a laboratory environment are not necessary or relevant to insomnia treatment.'
Peer Reviewer 7	Methods	Doc p. 10 line 42 …both investigators' summary assessment	Apostrophe added. Thank you.
TEP Reviewer 8	Methods	In Table B and Table 4, the remission criterion for PSQI is incorrectly stated as a score <5. Remission is indicated by a score less than or equal to 5. The text seems to get this correct, so hopefully the analyses are not affected.	Thank you. We have corrected this in Tables B and 4: 'Remitter - total score less than or equal to 5 at endpoint'
TEP Reviewer 8	Methods	I would not characterize sleep efficiency as a more 'comprehensive' sleep measure (ES-2). In fact, it is both a derivative measure and a ratio, which could arguably lead to it being a less desirable outcome.	Deleted 'comprehensive' when describing sleep efficiency.
TEP Reviewer 8	Methods	It was not clear to me whether the overall strength of evidence was based on a quantitative tallying of criteria, or a judgment on the part of the reviewers.	Revised text: 'Based on these factors, the overall strength of evidence for each outcome was judged as…'
TEP Reviewer 9	Methods	There is no mention of OTC anti-histaminergic agents in this section, or anywhere in the report. Given that these agents are, after alcohol, the most commonly employed chemical interventions for insomnia, this omission seems curious. One assumes this was deliberate, but some acknowledgment/explanation of the omission, at least, seems in order.	We included any treatment if it met our eligibility criteria. We did not identify eligible trials of antihistaminergic drugs or alcohol.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 9	Methods	A significant issue regarding extracted data first arises here and runs its course throughout. It appears (although is not explicitly states, that I can find) that only subjective, patientreported data is employed. Mention is made that this seems most relevant, as it is patient- centered. However, that is not to say that objective, polysomnographic data is not relevant. It is by no means entirely clear what the optimal metrics are for assessment (subjective or objective sleep data, daytime functional assessment, etc). This omission, and, more importantly, the failure to clarify the methodology and address this issue seems a significant short-coming.	Revised text: 'We included studies that report subjective outcomes. Polysomography outcomes are not patient-centered and trials reporting only these outcomes were excluded. Providers use history and patient report to diagnose insomnia disorder and assess patient opinion regarding treatment. Providers are more likely to value a patients' perspective of improvement based upon their typical sleep routine. Sleep parameters obtained in a laboratory environment are not necessary or relevant to insomnia treatment.'
TEP Reviewer 9	Methods	ES6. Table B. PSQI – LOWER scores indicate better sleep	Thank you. Corrected in Tables B and 4.
TEP Reviewer 9	Methods	P8. Criteria for inclusion – Para 1. It is stated that patients with 'certain comorbidities' were included, but then goes on to seemingly imply that comorbidities such as depression etc. are excluded?? This is an important issue which deserves considerably more clarification, if not an exegesis of which comorbidities are associated with which studies and outcomes.	Added clarification: 'Coexisting diseases are common among patients with sleep problems, so we included studies that enrolled participants with certain comorbidities. Other medical or mental health conditions (e.g., pregnancy, menopause, major depressive disorder, bipolar disorder, post-traumatic stress disorder, fibromyalgia, rheumatoid arthritis, Parkinson's disease, etc.) may explain insomnia symptoms, and therefore trials enrolling these subgroups do not meet diagnostic criteria for insomnia disorder and are excluded.'
TEP Reviewer 9	Methods	The exclusion of pharmacological duration of studies less than 4 weeks duration is questionable. Although the point is made that we are dealing predominantly with a chronic problem, there are data which suggest short-term pharmacotherapy, accompanied by CBT-I, with fairly rapid discontinuation of meds, may produce the best outcomes. It is open to debate whether two week and four week durations are really so different in this context. All of that said, two week studies were included in the previous meta-analyses. I assume that comparison of these data to the present analysis yielded no significant difference in strength of evidence, though I am not aware of an explicit statement to that effect.	Revised text to clarify: 'Insomnia disorder is a chronic condition, so a study duration of at least 4-weeks was required for eligibility.' A study with CBT-I and short-term drug treatment would be eligible if outcomes were measured at four weeks or beyond even if the drug component of the treatment had been discontinued. Strength of evidence assessment methods have changed considerably since the last review. We added text to the discussion to discuss the differences/similarities of our review to previous reviews:

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 10	Methods	Methods generally sound. A few comments:- the overall risk of bias assessments reflected 'investigators' subjective assessment'. I'm not opposed to this, but it would be useful to know whether any guidance was provided, whether these were investigators experienced in use of Cochrane tool, and how often the two investigators disagreed. Again, probably not a big deal, but it could be if there were inexperienced investigators and there was a lot of disagreement - we struggle with these assessments and I could imagine significant variation in subjective assessments without some clear guidance	We elaborated on the specific risk of bias assessments and included investigator guidance in the appendices. Revised text: 'Two investigators independently assessed risk of bias for eligible RCTs using an assessment tool developed for this project (Appendix B). Investigators assessed several types of bias including selection bias (method of randomization, groups similarity at baseline, allocation concealment), performance bias (blinding of provider and recipient, intervention definition – theory based, manualized, fidelity to treatment), detection bias (outcome assessors blinded, instruments validated and reliable, clinical significance of outcomes, co-interventions avoided or similar, correction for multiple comparisons, power – if pooling not possible), attrition bias (extent of attrition, reasons for incomplete data provided, incomplete data handled appropriately), reporting bias (select group of outcomes reported, select analysis conducted), and other sources of bias. Certain items (such as adequacy of intervention definition and implementation) were especially necessary to adequately capture all potential risk of bias associated with psychological interventions.' There was often disagreement in overall assessments. We only excluded a study from analysis when both investigators assessed the trial as high risk of bias.
Peer Reviewer 10	Methods	not sure what vote-counting method means in this context	Removed text. We did not end up using vote counting (assessing the number of trials showing effectiveness for a particular comparison-outcome combination over number of sufficiently powered trials).
Peer Reviewer 10	Methods	'based upon the number of limitations detected during risk of bias assessments' - yet above it sounded like the overall risk of bias assessment was judged subjectively. Please clarify.	Revised text to clarify: 'Based on study design and conduct of the individual studies making up the body of evidence for a particular comparison, study limitations were rated low, medium, or high based upon the number and magnitude of limitations detected during risk of bias assessments.'
Peer Reviewer 10	Methods	- page 8 - analytic framework - I would suggest revising the framework to more clearly show the relationship among the different outcome categories. Sleep outcomes like sleep latency should be proximal to daytime functioning and QOL. This should also be clarified in the relevant sections in intro and discussion.	Thank you. Great idea. Revised analytic framework.

Commentator & Affiliation	Section	Comment	Response
		Not crazy about the use of the term 'secondary' outcomes here. Often this term is used to describe less important (or at least non-primary outcomes). In this case it is being used to explain that these are the health consequences of sleep outcomes. It might be better to call these health outcomes.	Our thoughts in classifying certain outcomes as secondary were that these outcomes such as functioning, mood, and quality of life were secondary because they worked through improved sleep and were not direct consequences of the treatment. However, the terminology seemed confusing and undermining to several peer reviewers and we will revise the text to call these outcomes 'functioning, mood, and quality of life outcomes'.
Peer Reviewer 10	Methods	page 8 - lines 43-45 - After reading this it is not exactly clear whether or not you included studies which included patients with depression. Please clarify as this is a critical issue in determining applicability of this review.	Revised text to clarify: Studies were eligible if they specifically enrolled patients with 'mild depression' or if they enrolled patients with comorbidities not fully described. 'Coexisting diseases are common among patients with sleep problems, so we included studies that enrolled participants with comorbidities (sometimes called 'secondary insomnia') and trials enrolling pure subgroups of patients with certain conditions (i.e. anxiety, mild depression, non-cancer pain). Other medical or mental health conditions (e.g., pregnancy, menopause, major depressive disorder, bipolar disorder, post-traumatic stress disorder, fibromyalgia, rheumatoid arthritis, Parkinson's disease, etc.) may explain insomnia symptoms, and therefore trials enrolling these subgroups were excluded; it is not clear that these patients meet diagnostic criteria for insomnia disorder.'

Commentator & Affiliation	Section	Comment	Response
Public Reviewer 5 Richard Chapell, Assoc. Director HTACERCOREUS Outcomes Research, Merck Co. Inc.;For the Merck Insomnia Comment Team		Concern. Thank you for the opportunity to comment on the draft AHRQ report on management of insomnia disorder. We have found it to be well organized and informative. However we notice that the literature searches on which the report is based are already a full year old. We presume that these will be updated prior to finalizing the document.The timeliness of the report is especially important because in the time since the reports inception at least one new product has received FDA approval. Suvorexant marketed in the US under the brand name Belsomra was approved in August and will soon be on the market. To date two publications describing two clinical trials of suvorexant have been published. We have attached abstracts below for the convenience of your reviewers. We request that these reports be included in the systematic review in order to make it as up to date as possible.If for reasons of time and budget the EPC is unable to review the two publications we request that suvorexant be added to the reviews list of treatments marketed in the United States along with an explanation as to why the treatment is not included in the review.	We have updated the search and included the suvorexant trials.
Peer Reviewer 1	Results	In the discussion of zolpidem 'as needed' (Walsh, 2002), the detailed data are contained in Walsh et al., Sleep 2000;23(8):1-10. The zolpidem was supplied 3-5 times per week and contrasted with placebo. Sleep was better on the zolpidem nights as contrasted to placebo, but sleep was WORSE on other nights than the placebo group, indicating withdrawal insomnia. 'As needed' use made insomnia WORSE on nights the drug was omitted. The overall amount of sleep, combining nights when zolpidem was and was not taken, was NOT significantly greater than placebo. Since the overall effect was NOT a significant improvement, the 'as needed' use was overall ineffective, so this review misinterpreted the outcome. To consider the result only 'on nights when the medication was taken,' (page 85, lines 23-25) is to bias consideration towards only the favorable aspects of 'as needed' use, without considering the detrimental effect of drug on nights the drug is not taken. The reviewer does not know if the same problem was revealed in the other 'as needed' studies, but that should be examined.	We aimed to assess the efficacy of the active treatment when it was administered. However, we have added the information that when all nights (on and off zolpidem) were taken into consideration, there was not significant improvement versus placebo in sleep onset latency and other outcomes.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Results	On page 90, lines 47-48, the systematic review violated its guidelines to use randomized comparisons between treatments and placebo, by commenting on 'rebound' criteria which are comparisons of drug withdrawal with baseline lacking control for order or cross-over. Longitudinally, both the placebo group and the eszopiclone group in Ancoli-Israel et al. (2010) showed spontaneous remission over time, accounting for the lack of rebound. The correct comparison in a design with a randomized parallel placebo group is between the drug group and the placebo group after drug withdrawal. Figure 2 of that publication shows conclusively that sleep latency was longer (WORSE) in the eszopiclone group than in the placebo group for up to 28 days after the drug was withdrawn, which is a clinically significant harm. Also, total sleep time was distinctly worse in the drug group than the placebo group on the first night of drug withdrawal, showing a sudden deterioration which would be clinically disturbing. Had the objective actigraphic data from the study offered at www.clinicaltrials.gov been considered, the reviewers would have found a far less favorable picture of the drug than was presented with subjective data. The 'rebound' criteria seem to have pulled the wool over the eyes of reviewers who were trying to evaluate black sheep.	We have noted the rebound insomnia that occurred in eligible trials during the discontinuation period when it was reported. We have stated 'the percentage of participants with ISI total scores categorized as 'no insomnia' and 'subthreshold insomnia' declined in the eszopiclone group from 78% at week 12 when treatment was discontinued to 53% at week 16. A regression of sleep latency in the eszopiclone group to the level of the placebo group was also observed at day 28 after the drug was withdrawn.'
Peer Reviewer 1	Results	Joya et al. (J. Clin. Sleep Med., 2009) examined 36 hypnotics trials for information about infection, 13 of which lasted 1 month or longer. The majority provided infection data. Formal meta-analysis demonstrated a higher incidence of infection among drug groups than among randomized placebo groups. Three trials for eszopiclone and zolpidem had sufficient data to conclude that infection was significantly greater in the drug group in the single controlled trial by itself. Infection is currently listed in official prescribing information as a common risk of several hypnotics. Failure of the comparative effectiveness review to mention the word 'infection,' much less to seriously analyze this common risk of hypnotics, is a striking example of the review's failure to adequately examine and compare the harms of hypnotics and other treatments of insomnia disorder.	We've added results of observational studies of hypnotic use for insomnia to our results.
TEP Reviewer 2	Results	The overall results are clear and well-reported. Please see specific minor issues for typographical and syntax errors (listed under Clarity and usability).	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 3	Results	There is a considerable amount of detail in the results section. Characteristics of the study are clearly described. [specific comments re: ES results appear below]	Thank you.
Peer Reviewer 4	Results	All fine	Thank you.
TEP Reviewer 6	Results	The Results are presented in very technical terms and are unlikely to be understandable by the non-sleep specialist and not clinically useful to sleep clinicians. More meaningful unit of measurements should be use to describe the magnitude of improvements (i.e., reduction of 30 minutes of sleep latency or a reduction of 9 points on the ISI).	We have tried to express the magnitude of effect of treatments in terms of ISI scores or minutes.
		In the section comparing effectiveness of different interventions (p. 86), the Morin et al (1999) study and Sivertsen et al (20??) should also be included as these studies met inclusion criteria and specifically compared psychological and pharmacological therapies.	The Morin 1999 study was included in our comparative analysis; we did not formally extract the data in the draft because this study was included in the SR we extracted. We have now decided to use the RCTs instead of the SR for this comparison. The Sivertsen trial does not meet our inclusion criteria because it studies zopiclone, a drug that is not FDA approved for any indication.
TEP Reviewer 6	Results	This reviewer is in disagreement with some of the outcomes reported. For instance, it is not entirely clear how this review can conclude that doxepin improved global outcomes in older adults given the very limited number of studies available and that there is insufficient evidence on benzodiazepines for all outcomes and populations. This does not make much sense given the several hundred studies and systematic reviews/meta-analyses (Glass et al.; Holbrook et al.; Nowell et al.; Smith et al.; Gross et al., 2011) that have been published on the efficacy and adverse effects of benzodiazepines.	Two doxepin trials (n=494) met our inclusion criteria and reported global outcomes. Evidence for global outcomes was assessed as insufficient and evidence for certain sleep outcomes was assessed as low. We excluded most benzodiazepine trials for inadequate duration and/or no patient-reported outcomes.
TEP Reviewer 8	Results	Detail in REsults is appropriate. Studies are clearly described. The key messages are also clearly described. See comments regarding a couple of studies.	Thank you.
TEP Reviewer 8	Results	In Brief Behavioral Treatments for older adults, two of the references report on the same sample. Specifically, Germain, 2006 reported on a subset of participants and measures reported in Buysse, 2011 (this is indicated in the Methods section of the latter paper).	Thank you for bringing this to our attention; our analysis has be redone to avoid double-counting this trial.
TEP Reviewer 8	Results	I found the omission of Edinger's trials using a quasi-desensitization control from the 'efficacy' section to be odd. The very reason for developing this intervention was to	Thank you for bringing this to our attention; we were aware of the Edinger trials and were initially reluctant to pool them with trials using no treatment or waitlist

Commentator & Affiliation	Section	Comment	Response
		provide a more rigorous control than typically used in psychological intervention studies. Edinger and colleagues went to some lengths to ensure that this intervention was both plausible and lacking in efficacy. Thus, I would argue that it provides a higher standard of evidence for the true efficacy of CBT-I, and should not be relegated to a qualitative description in the comparative efficacy section.	controls. These are higher quality trials and we have incorporated their findings into our analysis. Unfortunately, data is sparse to separately analyze different types of control conditions; we did small separate analyses with overall pooling.
TEP Reviewer 8	Results	I would not use the term 'multicomponent behavioral therapy' to describe brief behavioral treatment. The word 'multicomponent' is widely used in association with cognitive behavioral treatment for insomnia.	We tried to use the terms multicomponent behavioral therapy and brief behavioral therapy to represent different treatments; however, we felt they had similar components and could be pooled to strengthen the available evidence.
TEP Reviewer 8	Results	I would be cautious about using the term 'combined treatment' for psychological + pharmacologic treatment. Specifically, the paper also talks about combinations of psychological interventions, so 'combined' may lack the necessary precision.	We believe the term 'combined' was more understandable than 'multimodal' and specified what the combinations were when using the term.
TEP Reviewer 8	Results	I would also consider alternatives to 'modest' sleep improvements (e.g., ES-9). We don't really know what magnitude of sleep improvement is clinically or biologically important, nor can we assume a linear function. In other words, a change in sleep latency of 15 minutes could actually be quite important for insomnia patients. Conversely, a reduction from 30 to 15 minutes is a 50% reduction numerically, but probably is probably not literally 'half as bad.' More neutral terms referring more directly to the magnitude of change may be more desirable.	Removed 'modest' and reported only minutes: 'Melatonin PR showed decrease in sleep onset latency of 6 minutes. Low-strength evidence shows that ramelteon did not improve sleep outcomes when compared with placebo.'
TEP Reviewer 8	Results	It was not clear to me whether the overall strength of evidence was based on a quantitative tallying of criteria, or a judgment on the part of the reviewers.	Please see the Methods section of the full report for how strength of evidence is assessed. While these methods are fairly transparent in determining the individual components of SOE, judgment plays a part.
TEP Reviewer 8	Results	CBT-I vs medication comparisons are always challenging to analyze, but I think the statement on ES-20 may be too strong. There has not yet been a study that compares CBT-I to true long-term pharmacotherapy. Existing trials used short-term pharmacotherapy and measured longer-term outcomes.	Thank you. We have decided to reanalyze the original trials as opposed to relying on the previously conducted systematic review; results from our analysis were much less conclusive. Unfortunately the evidence was insufficient to draw conclusions regarding comparative effectiveness.
TEP Reviewer 8	Results	The study inclusion criterion of 4 weeks for medications is understandable, but it does lead to the exclusion of many other studies. This may be a key point to emphasize in presenting the results: Many investigators/clinicians may think, perhaps rightly, that the actual number of published, reasonable quality studies is higher than presented here.	Agreed, we have tried to emphasize the short-term nature of the included trials and have noted that FDA approval specifies short-term use for many of these drugs.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 9	Results	Some explanations for the less well-informed reader would be appropriate in certain places. Here, for example, it might be noted that the instructions for SR and SC inherently and intentionally limit TST. Hence, it should be no surprise that increases in TST for certain specific CBT interventions are minimal.	We have updated Table 2 and Table A to better describe these interventions. We have also added text to the results section explaining this outcome: 'Since sleep restriction limits time in bed, it is to be expected that total sleep time would not differ significantly between sleep restriction and comparison groups.'
TEP Reviewer 9	Results	Throughout the results section, the issue of follow-up duration seems to be treated unevenly. In Figures 7 & 8, for example, the data are identified as 'at follow-up' but I can find no indication of length of follow-up, whereas in Tables 10-11, length of follow-up is clearly specified. Perhaps there is a reason but this issue requires checking throughout.	We divided outcome measurement timing into two categories and report outcomes measured between study initiation and six months as efficacy and comparative effectiveness. We report outcomes measured at six months and over as sustained efficacy.
TEP Reviewer 9	Results	[Pharmacologic Interventions] Key points. Here, again, some guidance to the reader may be in order. There is no acknowledgment in the presentation of data that one of the most important differences among these medications is half-life. One would hardly expect a medication with a 1 hour half-life to have much effect on WASO, for example. While the data, for the most part, speaks for itself in this respect, I still believe that some explanation would help the more naïve reader.	We have noted that the drugs have different half-lives (Introduction: 'Medications of varying half-lives…). Patients and providers specifically want to know the effects of medications on various sleep parameters. We have added text describe how specific drugs or formulations can be used to target specific sleep problems (Introduction: 'Drugs can be specifically formulated…'). Additionally, a table has been added to Appendix E that describes specific drug half-lives.
TEP Reviewer#9	Results	In addition, throughout this section, there also seems to be inconsistency in which evidence tables are included and which are not. I see no consistent pattern to this.	We tried to include evidence tables when at least two trials were available for a certain comparison.
TEP Reviewer 9	Results	P.35. Figure 23. 7.2.1 The table shows an improvement of 43m. On TST with p< .03. The text says no significant differences for SC on TST. Please check.	The text and figures for stimulus control have been updated with new data and are now consistent.
TEP Reviewer 9	Results	P. 40. Bright light therapy. It would be helpful to the reader to know the timing of administration of light in these studies	We added text to describe the timing of bright light therapy: 'Friedman et al. randomized 61 older adults to bright (~4,000 lux) or dim light in the morning or evening and reports on 51 completers.'
TEP Reviewer 9	Results	P. 41 Bullet 1. Should the next to last sentence read 'Low strength evidence shows that zolpidem improved sleep onset latency BUT had higher adverse effects.?' It is unclear as stands.	We have amended this statement for clarity.
TEP Reviewer 9	Results	P. 46. Sleep outcomes. Para 1. Final sentence. Krystal reported significant improvement on SOL, WASO, TST and SE at all points. Please check.	We have confirmed this statement.
Peer Reviewer 10	Results	'was more robust' - than what?'	This key point has been changed and no longer describes conclusions as 'more robust'.
Peer Reviewer 10	Results	key points - no mention of secondary outcomes - need to say whether or not you found anything as these are the outcomes we are actually trying to change in treating	We selected outcomes that would best enable analysis of the available literature and outcomes considered important to providers. Functioning, mood,

Commentator & Affiliation	Section	Comment	Response
		insomnia in the first place	and quality of life outcomes are important measures of treatment impact. Unfortunately the same outcomes are not often reported within same comparisons making analysis of this wide variety of outcomes and instruments very limited. We chose to prioritize global and sleep outcomes for this reason.
Peer Reviewer #10	Results	-ES-16 - lines 12-13 - sentence 'only adverse effects'.. awkward phrasing - there are several instances of this - draft will just need thorough copy-editing before finalizing. Es-20 line 56 is another example	Thank you. Our draft will go through extensive copy editing prior to posting.
Peer Reviewer 10	Results	- section on CBT - since this is rated as having highest strength evidence, it is all the more important to describe what CBT is and what these studies actually did.	We have updated Tables A and 2 to better explain what CBT-I entails.
Peer Reviewer 10	Results	CBT and applicability - this seems an important issue - these are tiny studies. Any info re: ratio of eligible to enrolled patients?	This is an important issue; we did not extract these data as this was beyond the scope and scale of this review.
Peer Reviewer 10	Results	- CBT and study methods - these studies are, perhaps, more prone to bias 2/2 lack of blinding than the drug studies which are placebo controlled. Many of the outcomes are subjective assessments. I would assume, for many of the CBT studies, that patients knew they were getting an intervention. Does this factor in to the strength of evidence ratings for this rx?	Yes, risk of bias is one key component of SOE. It is much more difficult for many behavioral studies to achieve an overall risk of bias assessment when compared to double-blinded placebo controlled drug trials. However, a few trials do attempt double blinding with sham treatments of similar hours.
Peer Reviewer 10	Results	- CBT and strength of evidence - in looking at the SOE tables - it is interesting that CBT garnered moderate SOE with about 100 patients total while some other interventions met same criteria and had many more patients (though fewer studies).	CBT trials may have had small numbers with respect to individual instruments and how global outcomes were measured, but the overall consistency across instruments (encompassing far more patients) provided more confidence in the estimates. CBT trials that provided data sufficient for pooling enrolled 1382 participants, far more than any other psychological intervention. Drug trials rarely reported global outcomes, so we are less confident about the few times they are reported.
Peer Reviewer 10	Results	- secondary outcomes - again, these are barely mentioned but are probably the most important outcomes. P 23 - several studies but no information.	Secondary outcomes were not reported consistently within comparisons; the scarcity of data on these outcomes within a comparison combined with their indirectness guarantees the data will be insufficient. We therefore relied heavily on global outcomes that incorporate various aspects of functioning and distress.
Peer Reviewer 10	Results	- trazadone - this is striking. One of the most commonly used drugs for insomnia. No evidence. Important finding - would say more in discussion. Saw a 2008 review in JGIM (Buscemi) which did include a large RCT comparing	No trial evaluating trazodone versus placebo met our inclusion criteria. We identified one trial included in the previous AHRQ systematic review. This trial (Walsh 1998) was excluded because it did not meet our

Commentator & Affiliation	Section	Comment	Response
		trazadone to zolpidem but don't see it included here - not sure if perhaps it met exclusion criteria for some reason.	specified duration of four weeks; that trial's duration was 14 days.
Public Reviewer 5 Richard Chapell, Assoc. Director HTACERCOREUS Outcomes Research, Merck Co. Inc.;For the Merck Insomnia Comment Team		Orexin receptor antagonism for treatment of insomnia. A randomized clinical trial of suvorexant, William Joseph Herring Ellen S. Snyder Kerry S. Budd Jill E. Hutzelmann Duane B. Snavely Kenneth Liu Christopher Lines Thomas Roth and David Michelson Safety and efficacy of suvorexant during 1year treatment of insomnia with subsequent abrupt treatment discontinuation a phase 3 randomised doubleblind placebo controlled trialDavid Michelson a Ellen Snyder a Erin Paradis a Mary ChenganLiu a Duane B Snavely a Jill Hutzelmann a James K Walsh b Andrew D Krystal c Ruth M Benca d Martin Cohn e Christopher Lines a Thomas Roth f W Joseph Herring	Thank you for the references. We have included the eligible suvorexant trials and have used the journal publication versions.
Peer Reviewer 1	Discussion	The overall approach to meta-analysis seems unusual and disturbing. The treatments (whether CBTI or pharmacologic) have been considered by type of treatment, and then fragmented within each treatment (e.g., pattern of administration within each drug), so that no large meta-analysis of many trials, whether of CBTI or pharmacologic agents (or subgroups, e.g., benzodiazepine agonists) is provided. Overall summary meta-analyses which integrate classes of treatment is missing.	Meta-analysis requires similarities in populations, comparisons, and outcomes. We felt our level of aggregation was scientifically justified and further aggregation would provide information that is not valuable to making treatment decisions.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Discussion	What is the basis for concluding that benzodiazepines and agonists (especially 'nonbenzodiazepines') are 'safe and effective' as stated in the Abstract conclusion? For example, majority of categories in Table E and Table F showed 'not significant' or 'insufficient evidence' for a response in total sleep time. It is important to emphasize when the doses said to have evidence for response in total sleep time (e.g., zolpidem 10 mg. and eszopiclone 3 mg.) were generally above the starting doses recommended by FDA, at least for women (who receive the majority of insomnia treatments). Evidently, the FDA does not consider those doses safe. It appears that the comparative effectiveness review's conclusion that nonbenzodiazepines are 'safe and effective' is not justified by the evidence summarized.	Our abstract no longer says that these medications are safe and effective. We have added an appendix outlining the recommended dosage for each of the evaluated drugs and added a section on harms from longer-term, non-randomized, studies. We have also added text to our discussion section with caveats regarding applicability of dosages studied in the eligible trials: 'Few pharmaceutical trials measured and reported global outcomes; low-strength evidence suggests that eszopiclone, 2 and 3 mg, and zolpidem, 10 mg 'as needed' improve global outcomes. It is important to note that recommended dosage of zolpidem is currently only 5 mg for women, and that these trials we composed predominantly of women.' Suvorexant dosages studied were higher than the recommended initial dosage.
TEP Reviewer 2	Discussion	The findings are clearly stated and well-summarized. The limitations are also adequately described. In my opinion, the future research section could be enhanced with a discussion of specific methodological shortcomings in the studies reviewed that could be improved in future research. This is mentioned above (see my comments in methods). Without specific guidance, it is difficult for investigators to understand what aspects of study design need to be improved. Moreover, it would also be more difficult to justify this in grant applications, if additional resources are needed to implement the additional level of methodological rigor (e.g., increased staff to ensure blinding). Being specific here, would enable investigators to point to a specific recommendations that would be helpful in justification of methodology and resources in grant applications.	We've added specific recommendations to the future research needs section: 'More rigorously conducted trials of psychological treatments (i.e., double blinded with sham treatments instead of waitlist controls; ensure adequate power for type of trial [efficacy, equivalence, superiority], test theory-based treatments with sufficient treatment definitions [i.e., manuals, protocols], conduct training and fidelity checks to ensure that treatments are delivered as intended, select the most appropriate outcomes instruments and operationalization of those instrument scores [i.e., remission or response when established is more indicative of efficacy than a mean change in scores], use appropriate statistical techniques [comparisons across groups, correction for multiple comparisons when appropriate, power calculations], documentation of study withdrawals by group with reason, etc.)' We have also recommended longer-term trials to assess long-term adverse effects and harms and that data on harms and withdrawals be better documented.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 2	Discussion	One provocative idea to consider including for future research: Would the authors consider suggesting research focused on establishing an index of benefits-to-harms for each treatment? This would allow for a more standardized comparison across different types treatments (e.g., psychological, pharmacological, CAM). For example, a pharmacological treatment that has moderate efficacy and moderate risks, might have a neutral index while a CAM treatment that has small efficacy but very minimal risk might have a slightly positive index. Such an index would be useful for patients to discuss with clinicians when deciding which treatment is appropriate.	Thank you for the idea. We have included conceptual research as part of our future research needs section. However, our focus was first on establishing clinically meaningful changes in global and sleep outcomes. This is likely a prerequisite to the development of an index. Additionally, some global outcomes do assess components of both benefit and harm.
TEP Reviewer 2	Discussion	Furthermore, in the discussion or conclusions, some commentary on specific ways to improve strength of evidence could improve the overall methodological rigor as the field moves forward.	We've added specific recommendations to the future research needs section (these would strengthen the evidence base): 'More rigorously conducted trials of psychological treatments (i.e., double blinded with sham treatments instead of waitlist controls; ensure adequate power for type of trial [efficacy, equivalence, superiority], test theory-based treatments with sufficient treatment definitions [i.e., manuals, protocols], conduct training and fidelity checks to ensure that treatments are delivered as intended, select the most appropriate outcomes instruments and operationalization of those instrument scores [i.e., remission or response when established is more indicative of efficacy than a mean change in scores], use appropriate statistical techniques [comparisons across groups, correction for multiple comparisons when appropriate, power calculations], documentation of study withdrawals by group with reason, etc.)'
TEP Reviewer 3	Discussion	The implications of the major findings are clearly stated. Limitations are described.	Thank you.
TEP Reviewer 3	Discussion	Page 91 Table 30- future research needs Since Espie's internet CBTi showed such impressive sleep results (per Page 20 of report), should a line or two be added about the possible use of this methodology in primary care?	We have comments about the wide range of CBT-I options available. SOE was rated as low. We report the findings from synthesizing the evidence; ideally, these results are interpreted and used to make recommendations or treatment decisions by guideline groups and/or clinicians.
TEP Reviewer 3	Discussion	Here are some other studies that might be important published recently: Greenblatt DJ1, Harmatz JS, Singh NN, Steinberg F, Roth T,	Thank you. These trials are not eligible for the current review because they do not enroll patients diagnosed with insomnia disorder. Both enroll healthy patients.

Commentator & Affiliation	Section	Comment	Response
		Harris SC, Kapil RP. Drugs Aging. 2014 Oct;31(10):731-6. doi: 10.1007/s40266-014-0211-3. Pharmacokinetics of zolpidem from sublingual zolpidem tartrate tablets in healthy elderly versus non-elderly subjects. Horoszok L1, Baleeiro T, D'Aniello F, Gropper S, Santos B, Guglietta A, Roth T. Hum Psychopharmacol. 2014 May;29(3):266-73. doi: 10.1002/hup.2395. A single-dose, randomized, double-blind, double dummy, placebo and positive-controlled, five-way cross-over study to assess the pharmacodynamic effects of lorediplon in a phase advance model of insomnia in healthy Caucasian adult male subjects.	
Peer Reviewer 4	Discussion	Limitations The impression one might receive from the results is that only recently licensed drugs work. Those of us in the field know that this is not true, but either the introduction or the discussion needs to point out that requirements for licensing have improved and changed for the better massively over the years, so that the studies expected now were not carried out at the time older drugs were licensed. This does not mean that eg benzodiazepines would not have met modern evidence standards, eg safe in long-term of treatment, improvement of daytime function etc, and it would be helpful to have this explained in limitations.	We've added text to our limitations to describe this issue: 'Trials meeting our inclusion criteria were predominantly for more recently FDA approved drugs. Few trials on benzodiazapines or antidepressants for insomnia disorder were identified. These were largely excluded because study duration was less than four weeks. Other systematic reviews aiming to assess the efficacy of very short duration of these medications are available.'
Peer Reviewer 4	Discussion	Future research section v good	Thank you.
TEP Reviewer 6	Discussion	Although this technical report is likely to be of significant value to investigators, it is not entirely clear how useful it will be to clinicians or to those who might want to derive Clinical Practice Guidelines. Perhaps the Discussion should relate the findings to insomnia morbidity and general lack of recognition and treatment due to limited resources.	We disagree that it will not be useful for clinicians or those deriving guidelines. We have not assessed the literature to determine the magnitude of insomnia morbidity or the lack of recognition or limited resources.
TEP Reviewer 6	Discussion	Recommendations for future research are useful and in line with the current state of the literature. Perhaps some recommendations about the need for more studies of the natural history of insomnia would be informative about the need and to intervene during the course of this sleep disorder.	Our review synthesized the literature studying the natural history of or screening for insomnia disorder, so this may be beyond the scope of our review.
TEP Reviewer 8	Discussion	Implications are clearly stated and appropriate. Future research section might be a bit more explicit about desirable characteristics of future pharmacotherapy trials. With regard to insomnia methodology, it might be useful to cite a	We added specific recommendations to the future research needs section: 'More rigorously conducted trials of psychological treatments (i.e., double blinded with sham treatments

Commentator & Affiliation	Section	Comment	Response
		previous SLEEP 2006 paper that made explicit recommendations for insomnia research studies that in some way dovetail with these recommendations.	instead of waitlist controls; ensure adequate power for type of trial [efficacy, equivalence, superiority], test theory-based treatments with sufficient treatment definitions [i.e., manuals, protocols], conduct training and fidelity checks to ensure that treatments are delivered as intended, select the most appropriate outcomes instruments and operationalization of those instrument scores [i.e., remission or response when established is more indicative of efficacy than a mean change in scores], use appropriate statistical techniques [comparisons across groups, correction for multiple comparisons when appropriate, power calculations], documentation of study withdrawals by group with reason, etc.)'
TEP Reviewer 9	Discussion	Consideration should be given to offering the reader some guidance, as noted above. This might include comments on expected duration of action of medications, and duration of benefits from medication and non-pharmacological treatments.	We sought to assess the longterm efficacy of interventions for insomnia disorder; few long term trials are available for psychological interventions; even fewer for pharmacologic. This is identified as a future research need in our discussion section (Future Research Needs Table): 'Additional long term trials to assess efficacy and comparative effectiveness of evidence-based interventions…'.
Peer Reviewer 10	Discussion	See above comments - would say more about relationship of outcomes, lack of 'secondary' outcomes data, maybe more about harms, might be worth highlighting things like lack of trazadone data.	We updated the discussion to include these points.
Peer Reviewer 1	General	Should this be a review of treatment of insomnia disorder (at least 3 months) or all insomnia? Why was that decision made?	To be as current as possible with terminology, our review aimed to address treatment of 'insomnia disorder' as defined by DSM-5. This is consistent with the publicly nominated topic of 'treatment of chronic insomnia' and the literature included largely used the older terminology of 'chronic insomnia'. Given that most trials included patients with insomnia with mean duration of several years, these trials likely fairly accurately reflect the new updated diagnostic criteria for insomnia disorder.
Peer Reviewer 1	General	What is the definition of short-term treatment durations. e.g, p. 32 lines 19-21: 'However, most RCTs had duration shorter than drug therapy is used in practice. It is possible that these RCTs did not capture rare serious adverse effects	We revised the statement to clarify our meaning: 'Eligible drug trials rarely lasted longer than 6 weeks. Individuals taking these medications for insomnia may stay on the medications for months to years.'

Commentator & Affiliation	Section	Comment	Response
		associated with long-term use.' What is the definition of the duration that drug therapy is used in practice, and according to what evidence? This reviewer suspects that the median hypnotic treatment has a duration of less than 4 weeks, and that may be true of CBTI as well. Note, page 33, line 8, 'Eligible drug trials rarely lasted longer than 6 weeks.' Even if one accepts that symptoms must persist for at least 4 weeks to diagnose 'insomnia disorder', it does not follow that treatment must persist for at least 4 weeks, and it often does not. Therefore, the target of trials of at least 4 weeks duration seems inappropriate.	Our review topic was the management of insomnia disorder, a condition that requires three months of symptoms for diagnosis. We did not believe that treatment durations of less than four weeks could cure the disease since these medications treat symptoms of insomnia disorder and have no effect once discontinued. Our stakeholders and Technical Expert Panel members agreed with this decision. While those trials should be included in a review regarding the management of insomnia symptoms; they offer little value to the management of insomnia disorder.
Peer Reviewer 1	General	This peer reviewer failed to note any statement that www.clinicaltrials.gov or manufacturer or FDA data were consulted to include additional evidence apart from trials published 'in peer reviewed journals' (page 44, line 10). It appears that trial registries, FDA databases, and manufacturer information may have been used as 'grey' literature to assess reporting bias (page 44, lines 44-51), but it does not appear that efficacy data or cogent inferences regarding bias were derived from these sources. Much important and relevant data can only be obtained from unpublished sources, particularly about unpublished adverse effects and daytime performance measures. In general, the quantity of evidence assembled seems insufficient and seems to understate what is available even from public sources, partly because of the artificial limit to studies of 4 weeks to 3 months (page 44, line 6).	We searched gray literature resources to identify additional trials, assess publication bias. Clinicaltrials.gov does not consistently have results posted for use in systematic reviews. We also expanded our search for data on adverse effects by searching for large, longterm, observational studies done to identify associations between drugs when used for insomnia disorder and rare and serious adverse effects. We did not limit our inclusion of studies to three months; we initially sought to determine short term efficacy and longterm efficacy. However, longterm data was rarely available; we did include these studies when available.
		Although the detailed tables noted moderate risk of bias in many of the studies, there was no overall assessment of the impact of bias on the results, and 'bias' was not mentioned in the abstract. Since it is known that hypnotics manufacturers bias publication towards their more successful trials (Giles J, Nature, 2006;2006;440:270-2. Mattila, Eur Neuropsychopharmacol. 2011;21:500-507), this compilation limited to published trials knowingly adopted the pro-manufacturer-bias of published work which it had partially recognized, as well as possible similar biases towards positive results from practitioners of other treatments.	The detailed tables describe the risk of bias inherent in the individual trial. The overall risk of bias in the body of evidence for each comparison-outcome is factored into the strength of evidence assessment. There is a component of strength of evidence devoted to reporting bias strength of evidence assessments appear in Appendix D for psychological trials, Appendix E for pharmaceutical trials, Appendix F for CAM trials, and Appendix G for combined/comparison of intervention types.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	General	As stated (page 5, line 17), daytime functioning is an important aspect of global outcomes of insomnia treatment. Many studies of hypnotics and other insomnia treatments have collected objective data on daytime functioning, for example, on objective performance measures. Hypnotics (and sleep restriction treatment) may impact certain critical daytime functions such as driving without crashes and ambulating without falls. No explanation was given why the extensive data available on objective daytime function was omitted, neglecting the relevant fact that such studies often show impairment of function rather than benefit.	Daytime functioning is an important outcome of insomnia treatment and is captured in the instruments we categorized as measuring global outcomes. Other measures of daytime functioning (i.e., fatigue) were considered 'secondary outcomes'. Our research team with input from a Technical Expert Panel conducted a process to determine which outcomes were of primary importance in evaluating treatment success or failure. We selected global outcomes and sleep outcomes and assessed the strength of evidence on these. We also captured data on secondary outcomes; however because these outcomes are more infrequently reported and there was wide heterogeneity among studies, evidence for these outcomes is likely insufficient.
Peer Reviewer 1	General	Clarity and Usability: Because the study did not incorporate data available from clinicaltrials.gov, the FDA, or manufacturers, and did not effectively correct for the known bias of the pharmaceutical literature, the conclusions are biased. Since this systematic review did not explicitly examine the harms of treatment, as was one of its goals, it does not inform policy.	Publication bias is incorporated into our strength of evidence.
TEP Reviewer 2	General	This is a very thorough and comprehensive review that clearly took considerable effort to aggregate the data and evaluate the evidence. In some ways, the review is reassuring in that the current treatments we are delivering through our standard of practice is supported by the evidence from the literature. In other ways, it illuminates areas for further improvement. Overall, the key questions are answered appropriately.	Thank you.
TEP Reviewer 2	General	Clarity and Usability: The overall clarity and usability is very good. The report is well-organized and main points are presented clearly. See my comments above for suggestions to enhance the impact of this report on policy or practice decisions.	Thank you.
TEP Reviewer 3	General	Yes, the report is clinically meaningful. Adults and older adult populations were examined separately and sufficient reason was provided to explain this.	Thank you.
TEP Reviewer 3	General	Clarity and Usability: The report is well structured and organized. Main points are clearly presented. Conclusions are generally well stated to inform policy decisions.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 3	General	Consider adding the results of the analysis about trazodone in the abstract and/or summary conclusion statements. Although there was only one well conducted study, this medication is highly prescribed that readers might be interested in this.	We included any trazodone studies that met our inclusion criteria. Readers interested in understanding the efficacy of trazodone for short time periods would likely need to use a different systematic review. We have mentioned this in our discussion: 'Trials meeting our inclusion criteria were predominantly for more recently FDA approved drugs. Few trials on benzodiazapines or antidepressants for insomnia disorder were identified. These were largely excluded because study duration was less than four weeks. Other systematic reviews aiming to assess the efficacy of very short duration of these medications are available.'
Peer Reviewer 4	General	Clinically meaningful for clinicians in USA. Definitions and key questions well described	Thank you.
Peer Reviewer 4	General	Clarity and Usability: Clear and organized. conclusions could be used in policy/practice but with clear statement of limitations including those in comments above	Thank you.
TEP Reviewer 6	General	Key questions are clearly defined and relevant to this review.	Thank you.
TEP Reviewer 6	General	Primary and secondary end points are appropriate to this literature and appropriate to the research questions.	Thank you.
TEP Reviewer 6	General	The construct of 'treatment acceptability' could also be incorporated as a secondary end point because no matter how effective a treatment is, if it is not acceptable to patients it will be of little clinical utility. The literature is quite clear about differences between patients' acceptability of psychological and pharmacological therapies for insomnia and this issue warrants some discussion.	Thank you for raising this important point; while we did not identify this outcome and extract it, we did not observe this as being commonly reported. We did attempt to capture withdrawals as a proxy for these types of issues. Unfortunately withdrawals are not always sufficiently reported in psychological trials.
TEP Reviewer 6	General	Clarity and Usability: The reports is generally well organized.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 6	General	Throughout the manuscript the authors refer interchangeably to multicomponent behavioral therapy and brief behavioral therapy (BBT-I) as being the same treatment. It is not! Multicomponent behavioral therapy, as correctly stated by the authors, refers to a combination of several clinical procedures (e.g., stimulus control, sleep restriction) but without a cognitive therapy component. BBT-I is a multicomponent behavioral therapy as it also includes more than one behavioral treatment without cognitive therapy. The descriptor 'Brief' has nothing to do with its content; it is simply referring to dosage or method of delivery and, as such, should not be used as a label to describe treatment content. BBTI could be contrasted with 'group therapy', 'telephone consultation', etc. My recommendation is to restrict comparative effectiveness within the 'psychological/behavioral domain' to CBT-I and other multicomponent behavioral therapies (without the cognitive component).	We do not mean to imply that multicomponent behavioral therapy and brief behavioral therapy (BBT-I) are the same; however because they both have more than one behavioral component – we combined them for analysis. We did not feel that this was inappropriate since appeared to be similar components. Pooling these studies also suggests a statistical homogeneity (I^2=0% for most pooled outcomes). If pooling these trials is inappropriate, these trials would be analyzed separately resulting in insufficient evidence.
TEP Reviewer 6	General	A similar comment could be made about zolpidem and zolpidem 'as needed'. These are not two different drugs; they simply differ with regard to dosage/delivery of treatment.	We classified zolpidem nightly and zolpidem as needed as different treatment regimens and therefore did not pool these data. We felt this was most scientifically appropriate.
TEP Reviewer 6	General	The statement 'most studies were conducted in the United States' (p. 15) is inaccurate. In fact, of the 16 CBT-I studies conducted with adults more than 60% were conducted in Canada, England, Scotland, Norway, and Sweden.	We removed that statement and documented the actual country of study conduct in each section.
Peer Reviewer 7	General	Well written, clear analysis. Key points approach is helpful to readers.	Thank you.
Peer Reviewer#7	General	Use of the 'minimum important difference' when possible will be greatly appreciated by clinicians who often struggle to understand what changes in scores on research scales will actually mean for their patients.	We agree.
Peer Reviewer 7	General	The overall low quality of the literature is disappointing but not unexpected. It didn't surprise me, and I am sure that it did not surprise the authors, that the literature back to 2004 didn't include many of the old stalwarts: trazodone and benzos generally.	We agree.
Peer Reviewer 7	General	I was surprised by the effectiveness of CBT, the low – moderate SOE notwithwithstanding.	The result is consistent with other reviews.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 8	General	The report is clinically meaningful. The target population is not as explicit as it might be. The key questions are clearly worded.	Thank you.
TEP Reviewer 8	General	Clarity and Usability: Yes, well-structured, organized, and clearly-presented.	Thank you.
TEP Reviewer 8	General	This comparative effectiveness review provides a comprehensive summary of psychological and pharmacologic treatments for insomnia. It relies on past systematic reviews and a new data extraction from the published literature. In general, I found the review to be clear, concise, well-written, and balanced. The criteria for study selection and the evaluation methods are clearly-described and appropriate. Beyond providing a useful summary and integration of available evidence, the paper points out important limitations, such as the infrequent use of global outcomes, the lack of attention to minimally important differences, the high placebo response rate, and methodological differences between psychological and behavioral studies.	Thank you.
TEP Reviewer 9	General	This is a comprehensive, evidence-based analysis of pharmacological and non-pharmacological treatments for insomnia. The study populations consist of adult individuals with chronic insomnia disorder. Meta-analyses are employed wherever multiple studies could be pooled. These analyses incorporate appropriate consideration of bias and heterogeneity of data. The key questions are straightforward and clinically appropriate.	Thank you.
Peer Reviewer 10	General	This is a large, comprehensive, generally methodologically sound report addressing an important and complex topic. I didn't see (but may have missed) an explict definition of intended audience, though I think it is probably self-evident that this is a topic with broad interest to primary care physicians and specialists alike. The one subtopic in which it may be useful to be more explicit about audience is when discussing cognitive behavorial therapy. Many primary care physicians will not know exactly what this entails and these interventions are not well described in the text. The key questions are appropriate.	Thank you. We agree. Our intended audience is primary care providers and we have enhanced our discussion of the psychological treatments in Tables A and 2.
Peer Reviewer 10	General	Clarity and Usability: Yes, generally well structured and clear except as discussed above.	Thank you.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer 1 Richard Chapell, Assoc. Director HTACERCOREUS Outcomes Research, Merck Co. Inc.;For the Merck Insomnia Comment Team	General	To whom it may concernThank you for the opportunity to comment on the draft AHRQ report on management of insomnia disorder. We have found it to be well organized and informative. However we notice that the literature searches on which the report is based are already a full year old. We presume that these will be updated prior to finalizing the document.The timeliness of the report is especially important because in the time since the reports inception at least one new product has received FDA approval. Suvorexant marketed in the US under the brand name Belsomra was approved in August and will soon be on the market. To date two publications describing two clinical trials of suvorexant have been published. We have attached abstracts below for the convenience of your reviewers. We request that these reports be included in the systematic review in order to make it as up to date as possible.If for reasons of time and budget the EPC is unable to review the two publications we request that suvorexant be added to the reviews list of treatments marketed in the United States along with an explanation as to why the treatment is not included in the review.	Yes, we have updated the literature search and have included the drug recently FDA approved for treatment of insomnia.
TEP Reviewer 2	Figures	page 8 - analytic framework - I would suggest revising the framework to more clearly show the relationship among the different outcome categories. Sleep outcomes like sleep latency should be proximal to daytime functioning and QOL. This should also be clarified in the relevant sections in intro and discussion.	We have revised the analytical framework.
TEP Reviewer 8	Appendix	Appendix C does not contain a table for risk of bias in behavioral studies of older adults. Was this intentional?	Thank you. This oversight has been corrected.
TEP Reviewer 3	Executive Summary	Page ES-2 Paragraph 2 'subjective measures are generally…because they are 'patient –centered.' It's not clear what this means as it relates to the advantage of subjective measures over objective measures. Perhaps something more like, 'because subjective measures account for patient perspective.'	Patient-centered outcomes include outcomes that are noticeable to patients and do not include only patient-report.
TEP Reviewer 3	Executive Summary	Page ES-2 Table A Cognitive Behavioral Therapy is repeated in row 6 and row 11	Thank you. We have corrected this error.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 3	Executive Summary	Page ES-6 Table B Row 4: Pittsburgh Sleep Quality Index, column 2 Measurement/instrument properties 'with higher scores indicating better sleep'; Please check whether higher scores indicate worse sleep if a score of <5 indicates remitted?	Thank you. We have corrected this error in the text.
TEP Reviewer 3	Executive Summary	Page ES-15 Lines 15-20 Bullet point numbers 3 and 4 #3'Data on antidepressants (trazodone and doxepin) – this appears to suggest that trazodone and doxepin were the only antidepressants that were examined. Consider inserting 'e.g.trazodone and doxepin'	We included trials of any drug that met our inclusion criteria. Unfortunately many did not due to their short durations.
TEP Reviewer 3	Executive Summary	Page ES-17 trazodone does not seem to be listed in Table E	We identified no eligible trials of trazodone with acceptable risk of bias.
TEP Reviewer 3	Executive Summary	Page ES-20 Key points Bullet point 1: 'A previous fair quality systematic review' (citation?).	We no longer refer to the 'previous fair quality systematic review in the executive summary. We decided to extract data on the comparative effectiveness of CBT-I versus medications instead of relying on a previous systematic review.
TEP Reviewer 3	Executive Summary	Line 9 'did not capture rare serious side effects associated with long term use' implies that there are side effects with long term use. Are there?	Revised sentence to say 'If rare serious adverse effects are associated with these medications, it is possible that the short term trials included in our review would not capture them.'
Peer Reviewer 7	Executive Summary	In the Executive Summary Introduction, I would not conflate 'sleep problems' with 'insomnia.' Sticking with the latter term is more accurate.	We used the term sleep problems to build up to the actual medical disorder. We felt an introduction was necessary.
TEP Reviewer 8	Executive Summary	ES-21, '…the studies were not good about recording…' sounds judgmental.	Rephrased 'Psychological interventions are noninvasive and assumed to be low-harm interventions, but few trials reported adverse effects or withdrawals and often reported withdrawals in the overall population as opposed to withdrawals by group.'
TEP Reviewer 9	Executive Summary	ES1. Para 1. Last sentence – add mental disorders	We have this sentence explaining 'Additionally, about half of insomnia cases coexist with a psychiatric diagnosis.'

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 9	Executive Summary	ES2. Para 1. There is no mention of OTC anti-histaminergic agents in this section, or anywhere in the report. Given that these agents are, after alcohol, the most commonly employed chemical interventions for insomnia, this omission seems curious. One assumes this was deliberate, but some acknowledgment/explanation of the omission, at least, seems in order.	The omission was not deliberate. We included any trial of any intervention if the trial met our inclusion criteria. I suspect these trials are few and may not have adequate duration.
TEP Reviewer 9	Executive Summary	ES4-5. Methods. A significant issue regarding extracted data first arises here and runs its course throughout. It appears (although is not explicitly states, that I can find) that only subjective, patient reported data is employed. Mention is made that this seems most relevant, as it is patient-centered. However, that is not to say that objective, polysomnographic data is not relevant. It is by no means entirely clear what the optimal metrics are for assessment (subjective or objective sleep data, daytime functional assessment, etc). This omission, and, more importantly, the failure to clarify the methodology and address this issue seems a significant short-coming.	Our topic refinement process identified patient reported outcomes as more important than polysomography outcomes. They are patient-centered and more applicable to practice.
TEP Reviewer 9	Executive Summary	ES6. Table B. PSQI – LOWER scores indicate better sleep	Corrected. Thank you.
TEP Reviewer 9	Executive Summary	ES10. Bullet 2. Sub-bullet 2. Some explanations for the less well-informed reader would be appropriate in certain places. Here, for example, it might be noted that the instructions for SR and SC inherently and intentionally limit TST. Hence, it should be no surprise that increases in TST for certain specific CBT interventions are minimal.	Enhanced explanations of these interventions in Table A.
TEP Reviewer 9	Executive Summary	ES14. Key points. Here, again, some guidance to the reader may be in order. There is no acknowledgment in the presentation of data that one of the most important differences among these medications is half-life. One would hardly expect a medication with a 1 hour half-life to have much effect on WASO, for example. While the data, for the most part, speaks for itself in this respect, I still believe that some explanation would help the more naïve reader.	Thank you. We felt that our terminology (extended release, short-acting) was preferable to reporting medication half-life.
TEP Reviewer 9	Executive Summary	Lines 48-49 ES-1: This is a bit confusing - AASM recs are supported 'by the highest quality evidence' but this review suggests evidence is low to moderate. Please clarify whether you mean that AASM claimed that their recs were based on highest quality evidence (and you disagree) or if this actually refers to your finding that, in general, there may have been stronger evidence for CBT than other treatments.	We have removed that statement.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 9	Executive Summary	Table A ES-2 - cognitive behavior rx listed twice. The definition just restates the term itself but I don't think will be informative to most non-psychologists who don't do CBT. What is a rx that has a cognitive component? Does this include the other things listed in the table like relaxation training and biofeedback?	Thank you. We corrected this error and have provided a better explanation in Table A.
TEP Reviewer 9	Executive Summary	ES4-5. Methods. A significant issue regarding extracted data first arises here and runs its course throughout. It appears (although is not explicitly states, that I can find) that only subjective, patient-reported data is employed. Mention is made that this seems most relevant, as it is patient-centered. However, that is not to say that objective, polysomnographic data is not relevant. It is by no means entirely clear what the optimal metrics are for assessment (subjective or objective sleep data, daytime functional assessment, etc). This omission and, more importantly, the failure to clarify the methodology and address this issue seems a significant short-coming.	Included outcomes are explicitly stated in the PICOTS section. Justification for these selections was provided in the Introduction.
Peer Reviewer 10	Executive Summary	ES-6 line 10 - not sure what vote-counting method means in this context	We have removed that terminology.
Peer Reviewer 10	Executive Summary	ES-6 line 52 - 'based upon the number of limitations detected during risk of bias assessments' - yet above it sounded like the overall risk of bias assessment was judged subjectively. Please clarify.	Detailed descriptions of risk of bias assessments are available in the full report.
TEP Reviewer 6	Structured Abstract	The Structured Abstract is not very informative for the reader who hasn't also read the paper. For example, such statement as '…CBT-I improves global outcomes by minimum important differences and modestly improves most sleep outcomes in the general adult population (low to moderate strength of evidence' does not provide any sense of content/end points or magnitude of improvements. A more informative abstract would include specific statements about commonly used unit of measurements for insomnia (i.e., sleep onset latency, wake after sleep onset, Insomnia Severity Index) and would provide actual means/absolute values (i.e., mean reduction of 9 points on the ISI or a reduction of 25 minutes on sleep latency or 50 minutes of wake time after sleep onset) for these end points. Given that the large majority of investigators and clinicians will not read the entire paper (and many may only read the Abstract), the authors should aim to communicate more explicit information about outcomes in this section.	We have revised the abstract.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 6	Structured Abstract	Page v. Abstract: In light of the somewhat daunting length and detail of this report, what is contained in the abstract is especially important. There is no consideration of duration of efficacy in the results section. Although the majority of studies are relatively short-term, some data is available regarding longer-term outcomes for both CBT interventions and pharmacotherapy. To the extent allowed by the data, consideration might be given to addressing this important issue here.	Few trials provided longterm results; we did not have sufficient data to analyze treatment by duration.
TEP Reviewer 6	Structured Abstract	In addition, there is no mention of outcomes for antidepressant medication in the abstract. Given the continued widespread use of such agents in general practice, a summary of this may be appropriate for the abstract.	We have added a line in the abstract regarding doxepin.