



BMJ. 2012; 345: e8343.

PMCID: PMC3544552

Published online 2012 Dec 17. doi: [10.1136/bmj.e8343](https://doi.org/10.1136/bmj.e8343)

## Effectiveness of non-benzodiazepine hypnotics in treatment of adult insomnia: meta-analysis of data submitted to the Food and Drug Administration

[Tania B Huedo-Medina](#), assistant professor,<sup>1</sup> [Irving Kirsch](#), associate director of program in placebo studies (PIPS), lecturer in medicine, professor of psychology,<sup>2,3</sup> [Jo Middlemass](#), research assistant,<sup>4</sup> [Markos Klonizakis](#), research fellow,<sup>4</sup> and [A Niroshan Siriwardena](#), professor of primary and prehospital health care<sup>4</sup>

<sup>1</sup>Department of Allied Health Sciences, University of Connecticut, 358 Mansfield Road U-2101, Storrs, CT 06269-2101, USA

<sup>2</sup>Harvard Medical School, Beth Israel Deaconess Medical Center, USA,

<sup>3</sup>School of Psychology, Plymouth University, Plymouth, UK

<sup>4</sup>Community and Health Research Unit, Lincoln School of Health and Social Care, University of Lincoln, Lincoln LN6 7TS, UK

Correspondence to: A N Siriwardena ; Email: [nsiriwardena@lincoln.ac.uk](mailto:nsiriwardena@lincoln.ac.uk)

Accepted 2012 Nov 30.

Copyright © Huedo-Medina et al 2012

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-commercial License, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non commercial and is otherwise in compliance with the license. See:

<http://creativecommons.org/licenses/by-nc/2.0/> and <http://creativecommons.org/licenses/by-nc/2.0/legalcode>.

### Abstract

**Objectives** To investigate the effectiveness of non-benzodiazepine hypnotics (Z drugs) and associated placebo responses in adults and to evaluate potential moderators of effectiveness in a dataset used to approve these drugs.

**Design** Systematic review and meta-analysis.

**Data source** US Food and Drug Administration (FDA).

**Study selection** Randomised double blind parallel placebo controlled trials of currently approved Z drugs (eszopiclone, zaleplon, and zolpidem).

**Data extraction** Change score from baseline to post-test for drug and placebo groups; drug efficacy analysed as the difference of both change scores. Weighted raw and standardised mean differences with their confidence intervals under random effects assumptions for polysomnographic and subjective sleep latency, as primary outcomes. Secondary outcomes included waking after sleep onset, number of awakenings, total sleep time, sleep efficiency, and subjective sleep quality. Weighted least square regression analysis was used to explain heterogeneity of drug effects.

**Data synthesis** 13 studies containing 65 separate drug-placebo comparisons by type of outcome, type of drug, and dose were included. Studies included 4378 participants from different countries and varying drug doses, lengths of treatment, and study years. Z drugs showed significant, albeit small, improvements (reductions) in our primary outcomes: polysomnographic sleep latency (weighted standardised mean difference, 95% confidence interval  $-0.57$  to  $-0.16$ ) and subjective sleep latency ( $-0.33$ ,  $-0.62$  to  $-0.04$ ) compared with placebo. Analyses of weighted mean raw differences showed that Z drugs decreased polysomnographic sleep latency by 22 minutes ( $-33$  to  $-11$  minutes) compared with placebo. Although no significant effects were found in secondary outcomes, there were insufficient studies reporting these outcomes to allow firm conclusions. Moderator analyses indicated that sleep latency was more likely to be reduced in studies published earlier, with larger drug doses, with longer duration of treatment, with a greater proportion of younger and/or female patients, and with zolpidem.

**Conclusion** Compared with placebo, Z drugs produce slight improvements in subjective and polysomnographic sleep latency, especially with larger doses and regardless of type of drug. Although the drug effect and the placebo response were rather small and of questionable clinical importance, the two together produced to a reasonably large clinical response.

### Introduction

Hypnotic drugs are often prescribed in primary care for insomnia.<sup>1</sup> Despite a reduction in prescribing of benzodiazepine hypnotics in the past decade, hypnotic use and costs remain high because of the introduction and increase in use of Z drugs,<sup>2</sup>

a group of non-benzodiazepine hypnotic drugs (including eszopiclone, zaleplon, and zolpidem), which act on the GABA ( $\gamma$  aminobutyric acid) receptor and are used in the treatment of insomnia. These are now the most commonly prescribed hypnotic agents worldwide. Prescriptions exceed costs of \$285m (£178m, €221m) in the United States<sup>3</sup> and £25m (€31m, \$40m) in the UK.<sup>4</sup> Although widely prescribed, Z drugs are not without risks. These include adverse cognitive effects (such as memory loss), psychomotor effects (such as falls, fractures, road traffic crashes), daytime fatigue, tolerance, addiction, and excess mortality<sup>5</sup> with no significant difference from benzodiazepines.<sup>6</sup> These established risks need to be weighed against the benefits.

Previous meta-analyses<sup>6 7 8 9</sup> of clinical trials of Z drugs have been prone to publication bias, such as unavailability of unpublished trials, selective or duplicate publication, and selective reporting of results in constituent studies.<sup>10 11</sup> An example of the distorting effects of these publishing practices was shown in a study by Mattila and colleagues.<sup>12</sup> This study compared European Public Assessment Reports of three drugs for insomnia to identify clinical trials that were performed between 1998 and 2007 for the purpose of registration of these drugs in the European Union. They found that the effect size of these drugs was 1.6 times larger when it was based on published data compared with the whole sample of studies, published and unpublished. They also found “remarkable inconsistencies in the reporting of the secondary end points, methods, results and, especially safety.” Different characteristics of included studies have not been examined as possible moderators of effects of Z drugs in previous meta-analyses.

One way of reducing the problem of publication bias is to analyse the effect of drugs that have been approved by governmental agencies with data derived from regulatory submissions.<sup>13</sup> Drug companies are required to provide information on all sponsored trials, published or not, when applying for new drug approvals.<sup>14</sup> Hence, the US Food and Drug Administration (FDA) files contain a complete dataset of published and unpublished trials up to the date of drug approval. We therefore undertook a meta-analysis of randomised placebo controlled parallel group studies of clinical effectiveness of Z drug hypnotics for insomnia in adults using only data provided to the FDA for drug approval.

Another concern with studies of hypnotics is the magnitude of the placebo response. We have considered the distinction between drug and placebo responses and drug and placebo effects.<sup>15 16</sup> A drug response is the change that occurs after administration of the drug. The effect of the drug is that portion of the response that is due to the drug’s chemical composition; it is the difference between the drug response and the response to placebo. A similar distinction can be made between placebo responses and placebo effects. The placebo response is the change that occurs after administration of a placebo. It includes such factors as improvement because of the natural course of the condition and regression toward the mean, as well as the placebo effect itself.

Previous studies have shown significant improvements in placebo arms in placebo controlled trials of hypnotic drugs.<sup>17 18</sup> Assessment of the magnitude of the placebo effect is important for understanding drug-placebo differences and their implications for clinical practice. For example, a small drug-placebo difference might lead to different treatment options if the drug and placebo are both effective rather than if neither are effective. Because change in the absence of placebo administration is rarely assessed in randomised controlled trials (and was not assessed in the trials contained in the FDA files), we could not assess the placebo effect. Therefore, we assessed changes in placebo groups, as well as those in drug groups, thus allowing us to establish the magnitude and significance of placebo responses, drug effects, and other variables that can moderate these outcomes.

## Methods

### Data source

For this systematic review we adhered to PRISMA guidelines.<sup>19 20</sup> We obtained data on all currently approved (non-benzodiazepine) Z drugs: eszopiclone, zaleplon, and zolpidem from the FDA website (see appendix 1).

### Study selection

The criteria for inclusion were randomised double blind controlled trials, recruitment of adults with primary insomnia (transient or chronic), an intervention comparing a Z drug with a placebo control, submission to the FDA before approval, sponsored by the manufacturer, and studies from any country or reported in any language (although we found only reports in English). Studies were excluded if they were crossover designs, included healthy patients with normal sleep, were single night studies with induced insomnia, or did not report any inference test or enough descriptive information (for instance, percentages or means and a variability measure for both groups and/or both time measures) as included studies were too heterogeneous and not large enough to estimate the missing information to calculate an effect size. We excluded crossover trials because of problems associated with reactivity, learning, carry over effects, and failure of blinding. Blinding failure is

more likely with crossover studies, leading to an enhanced placebo effect in the drug treatment arm, thereby increasing the likelihood of a false positive (type I) error. We did not include post-approval trials in our analysis because it is not possible to obtain access to all unpublished data for those trials.

### Data extraction and quality assessment

Two independent trained raters extracted information related to the study with high inter-rater reliability: mean Cohen's  $\kappa$  0.90, for categorical variables, and mean intraclass correlation  $r=0.92$  for continuous variables. Because of the nature of the FDA data, extractors were blind to researchers and institutions.

Methodological quality was assessed with the Jadad scale<sup>21, 22</sup> as adapted by Miller and colleagues<sup>23</sup> (see appendix 2). For each study, we extracted statistical data for drug and placebo. We also coded sample and study characteristics and included dimensions such as study identifier, year of publication, location/s of study (country and number of sites), and study duration. Data were extracted for two primary outcomes and eight secondary outcomes. The primary outcomes were polysomnographic and subjective sleep latency. The secondary outcomes were subjective and polysomnographic total sleep time, subjective and polysomnographic number of awakenings, subjective sleep quality, sleep efficiency, and subjective and polysomnographic time awake after sleep onset. Measured characteristics of participants included proportion of women, age, and sample type (outpatients, elderly, etc). Design characteristics included design type, recruitment method, intervention drug(s), treatment duration, and statistics reported. None of the trials reported race or ethnicity.

### Data synthesis and analysis

For both measures of sleep latency (polysomnographic and subjective) and the eight other sleep related outcomes, we calculated effect sizes as the mean difference between pre-test and post-test divided by the standard deviation (SD) of the pre-test value<sup>24</sup> for each group separately (that is, repeated measures effect sizes, correcting for sample size bias).<sup>25</sup> The standardised mean change in the placebo group was subtracted from that of the intervention group to evaluate the drug effect with respect to the placebo effect for each comparison (that is, effect size between groups adjusted for baseline). We calculated multiple effect sizes if the study reported more than one drug group or multiple outcomes. The latter were analysed separately to investigate main effects and moderators. For multiple dose studies, with the same or different drug but with the same control group, we controlled multi-treatment dependence by estimating the covariance among them<sup>26</sup> to analyse the effect of different drug and dose combinations. The sign of the effect size was set so that negative values signified a decrease of waking after sleep onset, sleep latency, and number of awakenings (all both polysomnographic and subjective). The sign of the effect size was positive for increases in polysomnographic and subjective total sleep time, polysomnographic sleep efficiency, and sleep quality.

We obtained repeated measures effects sizes for each group for those comparisons reporting means and SDs and used medians and interquartile ranges as the best approximation for those studies missing mean and SDs for drug and placebo. Sensitivity analysis was undertaken by comparing the main results, with and without those comparisons where median and interquartile range were used to obtain a standardised mean difference. Transformations were conducted to obtain effect sizes between groups for those cases where F test or P values were reported.<sup>27</sup> As 63% (41) of the comparisons did not report SDs of each group and those provided were largely heterogeneous, we have reported repeated measures results for only six studies (24 comparisons) with the most complete statistical data; two studies were not included in the final analysis as they did not report any inference test or variability measure either in their repeated or two groups measures. We have reported effect sizes in their raw metric for the same comparisons in parallel to facilitate clinical interpretation.

We examined the effect sizes with random effects models<sup>27, 28</sup> for weighted effect sizes and publication bias. Random effects models are more robustly generalisable as they assume variability not only within studies but also between studies, a relevant assumption when studies from different populations are integrated to account for sampling error and population variance. Moderation patterns were examined under mixed and fixed effects assumptions, but we have reported results only under the latter assumptions because of the lack of power to show any significant pattern under mixed effects models.<sup>29</sup> The homogeneity statistic,  $Q$ , determined whether each set of weighted mean effect sizes shared a common parametric effect size: a significant  $Q$  indicates a lack of homogeneity. To assess not only significance of the heterogeneity but also its size, we calculated the  $I^2$  index and its corresponding 95% confidence intervals<sup>30</sup> to determine and compare across outcomes the extend of the heterogeneity.  $I^2$  varies between 0 (homogeneous) and 100% (non-homogeneous), and if the confidence interval around  $I^2$  includes zero, the set of effect sizes is considered homogeneous.<sup>31</sup> We investigated possible asymmetries in the distribution of the effect sizes, which could indicate reporting bias, using the trim and fill technique,<sup>32</sup> Begg's strategy,<sup>33</sup> and Egger's test.<sup>34</sup> We analysed the total Jadad score as well as individual item scores to detect any possible bias effect on the overall results. Finally, we conducted sensitivity analyses with effect sizes with more than 2 SD from the

average effect size.

Moderator analyses were conducted for the main outcomes, polysomnographic and subjective sleep latency. To explain possible moderation of the variability of the overall effect sizes, we examined the relation between sample, methodological, or condition characteristics and magnitude of effect using a modified weighted least squares bivariate regression analyses with weights equivalent to the inverse of the variance for each effect size.<sup>27 27 35</sup> Because doses of different drugs are not equivalent, we also tested the drug by dose interaction. We used total score on the methodological quality scale as a moderator to analyse possible interaction with the final weighted effect sizes and have presented any significant pattern for either sleep latency or its subjective measure.

## Results

### Description of studies

In the data obtained from the FDA website, we identified 13 clinical trials comprising 4378 participants that examined 65 separate drug-placebo comparisons by type of outcome, type of drug, and dose and that met the inclusion criteria. Figure 1 shows the trial flow. Table 1 and appendix 3 provides descriptive features of the studies. Methodological quality of the studies ranged from 13 to 21 on the Jadad scale (mean 15.63, SD 1.8). Publication year and quality score were not significantly correlated ( $r=0.34$ ,  $P=0.28$ ).

Studies were conducted in North America (eight studies), North America and Europe (one study), South America (one study), or Australia (one study), with one study conducted entirely in Europe, and another study without location information. The mean duration of studies was 33.9 days (SD 33.3, range 14-180 days). Of the 4378 participants sampled, 61% were women, 61% were aged under 45, and the mean age was 49.6 (SD 13.3; range 38-72) years.

All 13 studies included comparisons of at least one of our primary outcomes. Ten studies (22 comparisons) assessed polysomnographic sleep latency and seven (11 comparisons) assessed subjective sleep latency. The eight remaining secondary outcomes appeared in fewer studies: four studies (seven comparisons) assessed subjective total sleep time, two (two comparisons) assessed total polysomnographic sleep time, four (six comparisons) assessed subjective number of awakenings, three (four comparisons) assessed polysomnographic number of awakenings, two (four comparisons) assessed subjective sleep quality, three (five comparisons) assessed sleep efficiency, three (three comparisons) assessed polysomnographic waking after sleep onset, and one (one comparison) assessed subjective waking after sleep onset.

Zolpidem was most commonly prescribed drug (eight studies); eszopiclone and zaleplon were assessed in three studies each (one study included both zolpidem and zaleplon). Zolpidem was prescribed in eight studies (15 comparisons) measuring polysomnographic sleep latency, zaleplon in three studies (six comparisons), and eszopiclone in only one study (one comparison). Only zolpidem and eszopiclone were used in studies measuring subjective sleep latency, in five (eight comparisons) and two (three comparisons) studies, respectively.

### Quantitative analyses

For our primary outcomes, analyses of standardised effect sizes showed significant but small to medium differences in polysomnographic (weighted standardised mean difference  $-0.36$ , 95% confidence interval  $-0.57$  to  $-0.16$ ) and subjective sleep latency ( $-0.33$ ,  $-0.62$  to  $-0.04$ ) for treatment versus control. There were significant effect sizes for the primary outcome (sleep latency) within groups separately for both placebo ( $-0.39$ ,  $-0.54$  to  $-0.23$  (for polysomnographic);  $-0.33$ ,  $-0.63$  to  $-0.03$  (for subjective)) and drug ( $-0.93$ ,  $-1.32$  to  $-0.54$  (polysomnographic);  $-0.67$ ,  $-1.30$  to  $-0.03$  (subjective)). Analyses of weighted mean raw differences indicated that drugs decreased sleep latency by 22 minutes ( $-33$  to  $-11$  minutes).

Tables 2 and 3 show standardised and raw effect sizes, respectively. Figures 2 and 3 show forest plots for polysomnographic and subjective sleep latency, respectively. Analysis of secondary study outcomes showed no significant drug effect. The lack of difference between groups for other sleep measures coupled with the fact that few reports included them meant there was insufficient evidence to show efficacy on these measures.

There was no evidence of asymmetry of the distribution of the effect sizes for sleep latency by the trim and fill technique, Begg's test<sup>33</sup> ( $P=0.88$  (polysomnographic);  $P=0.22$  (subjective)), or Egger's test<sup>34</sup> ( $P=0.34$  (polysomnographic);  $P=0.77$  (subjective)), which suggests that these results are not significantly affected by publication bias. One study was an outlier (eszopiclone study No 190-047, table 1 and appendix 3), with a large pooled effect size for sleep latency 0.46 (0.24 to 0.69) and  $>2$  SD from the overall weighted effect size. When we excluded this study, the pooled effect size for sleep latency was  $-0.54$  ( $-0.91$  to  $-0.15$ ). Sensitivity analysis showed no significant differences in overall efficacy (the overall effect size without the outlier was still significant, with a slightly lower reduction of subjective sleep latency) and the same patterns for

the moderator results adjusted for the two outlier comparisons provided by this study.

Every item was evaluated through bivariate weighted regression analysis under fixed and random effects assumptions to critically and robustly appraise any included study for risk of bias in attributing outcomes to the intervention and their possible effect on the overall efficacy, but none of the results was significant. Therefore, there was no evidence of any interaction between quality/risk of bias in the included studies and the final results.

### Moderator effects on sleep latency

The main outcomes, polysomnographic and subjective sleep latency, were the only measures with sufficient cases to permit detailed models for moderator analyses (table 4). Sleep latency was more likely to be reduced in studies published earlier, with larger drug doses, longer treatment duration, and samples that included a greater proportion of younger patients and/or female patients (table 4). Polysomnographic and subjective sleep latency were reduced when larger doses were used, regardless of type of drug. The interaction of dose by type of drug was not significant, and all drugs (zolpidem and zaleplon for polysomnographic and subjective sleep latency and eszopiclone and zolpidem for subjective sleep latency, the latter being significantly more effective in this particular outcome) showed a pattern of greater reductions in sleep latency with larger doses. Subjective sleep latency was more likely to be reduced in studies published earlier, or with greater numbers of younger patients or women included in the sample, and with zolpidem. These patterns were obtained under fixed effect meta-regression models and these held under mixed effects assumptions.

## Discussion

### Main findings

In this meta-analysis of Z drugs using data published on the FDA website, which are less likely to be affected by selection or reporting bias, we found significant reductions in polysomnographic and subjective sleep latency in both drug and placebo groups. The difference between drug and placebo was 22 minutes for polysomnographic sleep latency and seven minutes for subjective sleep latency. Although these reductions in sleep latency might have benefits, albeit short term, for quality of life, the effect sizes corresponding to these differences were  $-0.36$  and  $-0.33$ , both of which are conventionally considered to be small effects,<sup>36</sup> and well below the criterion for clinical significance (0.50) suggested by the National Institute for Health and Clinical Excellence (NICE) in their guidelines for the treatment of depression.<sup>37</sup>

There were insufficient data for other drug effect end points to allow a valid analysis. The large heterogeneity in sleep latency outcomes was mainly explained by larger doses needed to obtain a greater drug than placebo effect. Z drugs were more likely to be effective in reducing sleep latency in studies published earlier, those including more younger and/or female patients, and those using zolpidem. Significant placebo responses were present in polysomnographic and subjective sleep latency. There have been several previous meta-analyses of published data on Z drugs, although none included moderator analyses and all acknowledged publication bias.<sup>4 6 8 9 38</sup>

### Strengths and weaknesses

As in previous studies, we found that data submitted for licensing enabled detailed investigation of drug efficacy.<sup>13 39 40</sup> We included sponsored studies submitted to the FDA but did not assess whether they were subsequently published. Studies submitted to the FDA are required to report all data so are less likely to be affected by reporting bias.

Studies were subjected to the same methodological scrutiny and analytical rigour as meta-analyses of published studies. As in other meta-analyses, we did not include studies that did not report enough statistical data to calculate an effect size. Because of the small number of reports for some outcomes, and the heterogeneity of statistical data reported, we could not compare some studies directly or robustly impute missing data. There was insufficient information about sample setting characteristics, drug side effects, and other factors that might have explained heterogeneity to fully account for these. The entry criteria for studies varied, with some studies focusing just on sleep latency, particularly for shorter acting drugs such as zaleplon. This could have affected the capacity of some studies to identify effects other than on sleep latency. All the drugs are licensed for insomnia, and patients presenting for treatment have a range of symptoms, not just sleep latency, for which these drugs are commonly prescribed in general practice.

Another weakness in the present analysis is that all the trials were industry sponsored. Industry sponsorship has been shown to enhance the outcome of clinical trials.<sup>41</sup> Thus, although we were able to include published and unpublished studies, at least for the reports used to approve these drugs, we could not avoid sponsorship bias, and our results might therefore overestimate the drug effect. Unfortunately, eliminating both sources of bias simultaneously is difficult, if not impossible. Although clinical trials now need to be registered in advance to be published in major medical journals,<sup>42</sup> there is no

requirement that the results be submitted for publication, and many failed clinical trials or clinical trials with negative results go unpublished.<sup>10</sup> Furthermore, although many clinical trials are subject to mandatory reporting of results to the FDA, most are not, and for those that are, as many as 78% fail to comply with this requirement.<sup>42</sup> Because sponsorship bias is in the direction of greater effects for industry sponsored trials, our results might overestimate the effects of Z drug hypnotics for treating adult insomnia.

We found evidence of a significant placebo response for sleep latency. McCall and colleagues undertook a meta-analysis of sleep changes associated with placebo in published hypnotic clinical trials and found a clinically important and statistically significant placebo response for subjective sleep latency and total sleep time.<sup>17</sup> Belanger and colleagues undertook a meta-analysis of sleep changes in control groups of 34 hypnotic drug studies in which 23 used a pharmacological placebo, four a psychological placebo, and seven a waiting list. They found significant pre-post changes in the pharmacological placebo group on several sleep outcomes, both objectively or subjectively measured, suggesting that sleep measures might change significantly in response to a pharmacological placebo.<sup>18</sup>

The response to placebo is more than just the placebo effect. Just as the effect of a drug is estimated by the difference between the response to the drug and the response to a placebo, the placebo effect would be the difference between the placebo response and changes occurring without administration of a placebo. Belanger and colleagues assessed the response to placebo hypnotic drugs and compared it with sleep changes among patients placed on a waiting list.<sup>18</sup> Compared with those on a waiting list, there were significantly greater improvements in subjective sleep onset latency (19.55 min v 2.43 min), subjective total sleep time (31.13 min v 7.30 min), and objective total sleep time (18.27 min v 10.34 min) in the placebo group.<sup>18</sup> These data were based on comparisons between studies rather than comparisons within studies, and none of the trials in the FDA database included waiting list controls. Nevertheless, the results of Belanger and colleagues suggest that the placebo response observed in our meta-analysis was largely caused by a genuine placebo effect. Future clinical trials including both placebo and untreated (natural course) controls would be useful, as well as combining the results of studies using network meta-analysis.

### Meaning of the study

The response to a medical treatment consists of two components: a true drug effect and a non-specific placebo response, which includes the placebo effect, regression toward the mean, and improvement because of the natural course of the condition. For that reason, it is useful, both for current clinical practice and for future treatment development, to know the effect sizes for the placebo group as well as for the control group. For example, finding that both placebos and drugs are effective but that the drug is more effective than the placebo, suggests that placebo characteristics can be used to amplify effectiveness of a drug. Conversely, finding improvement only in drug arms indicates that the placebo effect is not an important component of treatment, whereas finding that both are equally effective, compared with waiting list controls, suggests that non-specific aspects of patient care might be having positive effects.

We found that both the drug effect and the placebo response were small and of questionable clinical importance. The two put together, however, lead to a reasonably large clinical response. Although the drug-placebo difference in objectively measured sleep latency was only 22 minutes, the response to the Z drugs, including both drug effect and the placebo effect components, was 42 minutes. Similarly, the effect size for the drug response was  $-0.93$  and that for the placebo response was  $-0.39$ , accounting for about half of the drug response.

Insomnia is a symptom defined disorder characterised by distress about perceived poor sleep or lack of sleep. Hence, subjective sleep latency might be as important as objective sleep latency in understanding the benefits of treatments for this condition. The response to Z drugs was 25 minutes shorter for subjectively perceived sleep latency, whereas the response to placebo was an improvement of 19 minutes. Thus the benefit of Z drugs in term of subjectively perceived sleep latency was only seven minutes and was not significant. However, this was based on only two comparisons. Effect sizes for subjective sleep latency were calculable for a larger number of trials and the drug-placebo difference ( $-0.33$ ) was small but significant, with the placebo response again accounting for about half of the drug response.

Taken together, these data suggest that the placebo response is a major contributor to the effectiveness of Z drugs. The remaining effect needs to be balanced against the harms associated with these drugs. The substantial proportion of the drug response accounted for by the placebo response indicates the importance of non-specific factors in the treatment of insomnia. As the placebo effect is a psychological phenomenon, these data suggest that increased attention should be directed at psychological interventions for insomnia.

### Unanswered questions and future research

FDA data could also provide further opportunities for studying effects of adverse effects with Z drugs (particularly as larger effect sizes were associated with higher drug doses), as well as examining issues of publication and reporting delays and bias. We did not look at adverse effects, which can pose significant risks,<sup>43</sup> leading to concerns about the widespread and sometimes inappropriate use of these drugs.<sup>44 45</sup>

### Conclusion

This study of FDA data shows that Z drugs improve objective and subjective sleep latency compared with placebo, particularly in younger and female patients. The size of this effect, however, is small and needs to be balanced with concerns about adverse effects, tolerance, and potential addiction. The placebo response accounted for about half of the drug response. This suggests that increased attention should be directed at psychological interventions for insomnia.

#### What is already known on this topic

Z drug hypnotics have short term benefits in the treatment of insomnia  
Their effectiveness has been questioned because of publication bias reported in previous meta-analyses, and little is known about the extent of the placebo response

#### What this study adds

Z drugs decreased subjective and polysomnographic sleep latency compared with placebo especially with larger doses and in younger or female patients and regardless of type of drug  
The drug effect and the placebo response were small and of questionable clinical importance, but the two together produced a reasonably large clinical response  
The FDA clinical trial data were free from publication bias but might have been subject to other forms of bias

### Web Extra. Extra material supplied by the author

Appendix 1: Details of how to obtain drug trial data from FDA

Appendix 2: Modified Jadad score

Appendix 3: Full details of studies included

### Notes

Contributors: ANS and IK had the original idea for the study. All authors were involved in the design of the review, developed the search strategy, performed the study selection, interpreted and discussed results, and contributed to the writing and review of the various drafts of the report. JM, MK, and ANS extracted data from included studies. TBH-M, IK, and ANS were involved in data analysis. ANS is guarantor.

Funding: This study was funded by the College of Social Science Research Fund at the University of Lincoln. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests: All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required.

Data sharing: No additional data available.

## Notes

Cite this as: *BMJ* 2012;345:e8343

## References

1. Siriwardena AN, Apekey T, Tilling M, Dyas JV, Middleton H, Orner R. General practitioners' preferences for managing insomnia and opportunities for reducing hypnotic prescribing. *J Eval Clin Pract* 2010;16:731-7. [PubMed: 20545802]
2. Siriwardena AN, Qureshi Z, Gibson S, Collier S, Latham M. GPs' attitudes to benzodiazepine and 'Z-drug' prescribing: a barrier to implementation of evidence and guidance on hypnotics. *Br J Gen Pract* 2006;56:964-7. [PMCID: PMC1934058] [PubMed: 17132386]
3. Chilcott LA, Shapiro CM. The socioeconomic impact of insomnia. An overview. *Pharmacoeconomics* 1996;10(suppl 1):1-14.
4. Dunder Y, Boland A, Strobl J, Dodd S, Haycox A, Bagust, et al. Newer hypnotic drugs for the short-term management of insomnia: a systematic review and economic evaluation. *Health Technol Assess* 2004;8:iii-x,1-125.
5. Kripke DF, Langer RD, Kline LE. Hypnotics' association with mortality or cancer: a matched cohort study. *BMJ Open* 2012;2:e000850. [PMCID: PMC3293137]
6. Glass J, Lanctot KL, Herrmann N, Sproule BA, Busto UE. Sedative hypnotics in older people with insomnia: meta-analysis of risks and benefits. *BMJ* 2005;331:1169. [PMCID: PMC1285093] [PubMed: 16284208]
7. Dunder Y, Dodd S, Strobl J, Boland A, Dickson R, Walley T. Comparative efficacy of newer hypnotic drugs for the short-term management of insomnia: a systematic review and meta-analysis. *Hum Psychopharmacol* 2004;19:305-22. [PubMed: 15252823]
8. Buscemi N, Vandermeer B, Friesen C, Bialy L, Tubman M, Ospina M, et al. The efficacy and safety of drug treatments for chronic insomnia in adults: a meta-analysis of RCTs. *J Gen Intern Med* 2007;22:1335-50. [PMCID: PMC2219774] [PubMed: 17619935]
9. Holbrook AM, Crowther R, Lotter A, Cheng C, King D. Meta-analysis of benzodiazepine use in the treatment of insomnia. *CMAJ* 2000;162:225-33. [PMCID: PMC1232276] [PubMed: 10674059]
10. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine--selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 2003;326:1171-3. [PMCID: PMC156459] [PubMed: 12775615]
11. Godlee F, Loder E. Missing clinical trial data: setting the record straight. *BMJ* 2010;341:c5641. [PubMed: 20940217]
12. Mattila T, Stoyanova V, Elferink A, Gispens-de Weid C, de Boer A, Wohlfarth T. Insomnia medication: do published studies reflect the complete picture of efficacy and safety? *Eur Neuropsychopharmacol* 2011;21:500-7. [PubMed: 21084176]
13. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* 2008;5:e45. [PMCID: PMC2253608] [PubMed: 18303940]
14. Food and Drug Administration Center for Drug Evaluation and Research. Guidance for Industry: Guideline for the format and content of the summary for new drug and antibiotic applications. FDA, 1987.
15. Fisher S, Lipman RS, Uhlenhuth EH, Rickels K, Park LC. Drug effects and initial severity of symptomatology. *Psychopharmacologia* 1965;7:57-60. [PubMed: 5318924]
16. Kirsch I, Sapirstein G. Listening to Prozac but hearing placebo: A meta-analysis of antidepressant medication. *Prevention and Treatment* 1998;1:2a.
17. McCall WV, D'Agostino R Jr, Dunn A. A meta-analysis of sleep changes associated with placebo in hypnotic clinical trials. *Sleep Med* 2003;4:57-62. [PubMed: 14592361]
18. Belanger L, Vallieres A, Ivers H, Moreau V, Lavigne G, Morin CM. Meta-analysis of sleep changes in control groups of insomnia treatment trials. *J Sleep Res* 2007;16:77-84. [PubMed: 17309766]
19. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535. [PMCID: PMC2714657] [PubMed: 19622551]

20. Cooper HM, Hedges LV, Valentine JC, eds. *The handbook of research synthesis and meta-analysis*. 2nd ed. Russell Sage, 2009.
21. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1-12. [PubMed: 8721797]
22. Olivo SA, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: a systematic review. *Phys Ther* 2008;88:156-75. [PubMed: 18073267]
23. Miller WR, Brown JM, Simpson TL, Handmaker NS, Bien TH, Luckie LF, et al. What works? A methodological analysis of the alcohol treatment outcome literature. In: Hester RK, Miller WR, eds. *Handbook of alcoholism treatment approaches: Effective alternatives*. 2nd ed. Allyn and Bacon, 1995:12-44.
24. Becker BJ. Synthesizing standardized mean-change measures. *Br J Math Stat Psychol* 1988;41:257-78.
25. Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. *J Educ Behav Stat* 1981;6:107-28.
26. Gleser LJ, Olkin I. Stochastically dependent effect sizes. In: Cooper HM, Hedges LV, eds. *The handbook of research synthesis*. 1st ed. Russell Sage Foundation, 1994:339-355.
27. Lipsey MW, Wilson DB. *Practical meta-analysis*. Sage, 2001.
28. Schmidt FL, Oh IS, Hayes TL. Fixed- versus random-effects models in meta-analysis: model properties and an empirical comparison of differences in results. *Br J Math Stat Psychol* 2009;62:97-128. [PubMed: 18001516]
29. Hedges LV, Pigott TD. The power of statistical tests for moderators in meta-analysis. *Psychol Methods* 2004;9:426-45. [PubMed: 15598097]
30. Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychol Methods* 2006;11:193-206. [PubMed: 16784338]
31. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539-58. [PubMed: 12111919]
32. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;56:455-63. [PubMed: 10877304]
33. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;50:1088-101. [PubMed: 7786990]
34. Egger M, Davey SG, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-34. [PMCID: PMC2127453] [PubMed: 9310563]
35. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychol Methods* 2002;7:105-25. [PubMed: 11928886]
36. Cohen J. A power primer. *Psychol Bull* 1992;112:155-9. [PubMed: 19565683]
37. National Institute for Health and Clinical Excellence. *Depression: management of depression in primary and secondary care*. NICE, 2004 (Clinical practice guideline No 23).
38. Nowell PD, Mazumdar S, Buysse DJ, Dew MA, Reynolds CF, III, Kupfer DJ. Benzodiazepines and zolpidem for chronic insomnia: a meta-analysis of treatment efficacy. *JAMA* 1997;278:2170-7. [PubMed: 9417012]
39. Hemminki E. Study of information submitted by drug companies to licensing authorities. *BMJ* 1980;280:833-6. [PMCID: PMC1601011] [PubMed: 7370687]
40. MacLean CH, Morton SC, Ofman JJ, Roth EA, Shekelle PG. How useful are unpublished data from the Food and Drug Administration in meta-analysis? *J Clin Epidemiol* 2003;56:44-51. [PubMed: 12589869]
41. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252-60. [PubMed: 18199864]
42. Prayle AP, Hurley MN, Smyth AR. Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study. *BMJ* 2012;344:d7373. [PubMed: 22214756]

43. Siriwardena AN, Qureshi MZ, Dyas JV, Middleton H, Orner R. Magic bullets for insomnia? Patients' use and experiences of newer (Z drugs) versus older (benzodiazepine) hypnotics for sleep problems in primary care. *Br J Gen Pract* 2008;58:417-22. [PMCID: PMC2418994] [PubMed: 18505619]

44. Department of Health. National Service Framework for Mental Health, Modern Standards and Service Models. DOH, 2000.

45. National Institute for Health and Clinical Excellence. Guidance on the use of zaleplon, zolpidem and zopiclone for the short term management of insomnia. NICE, 2004 (Technology Appraisal Guidance 77).

### Figures and Tables

---

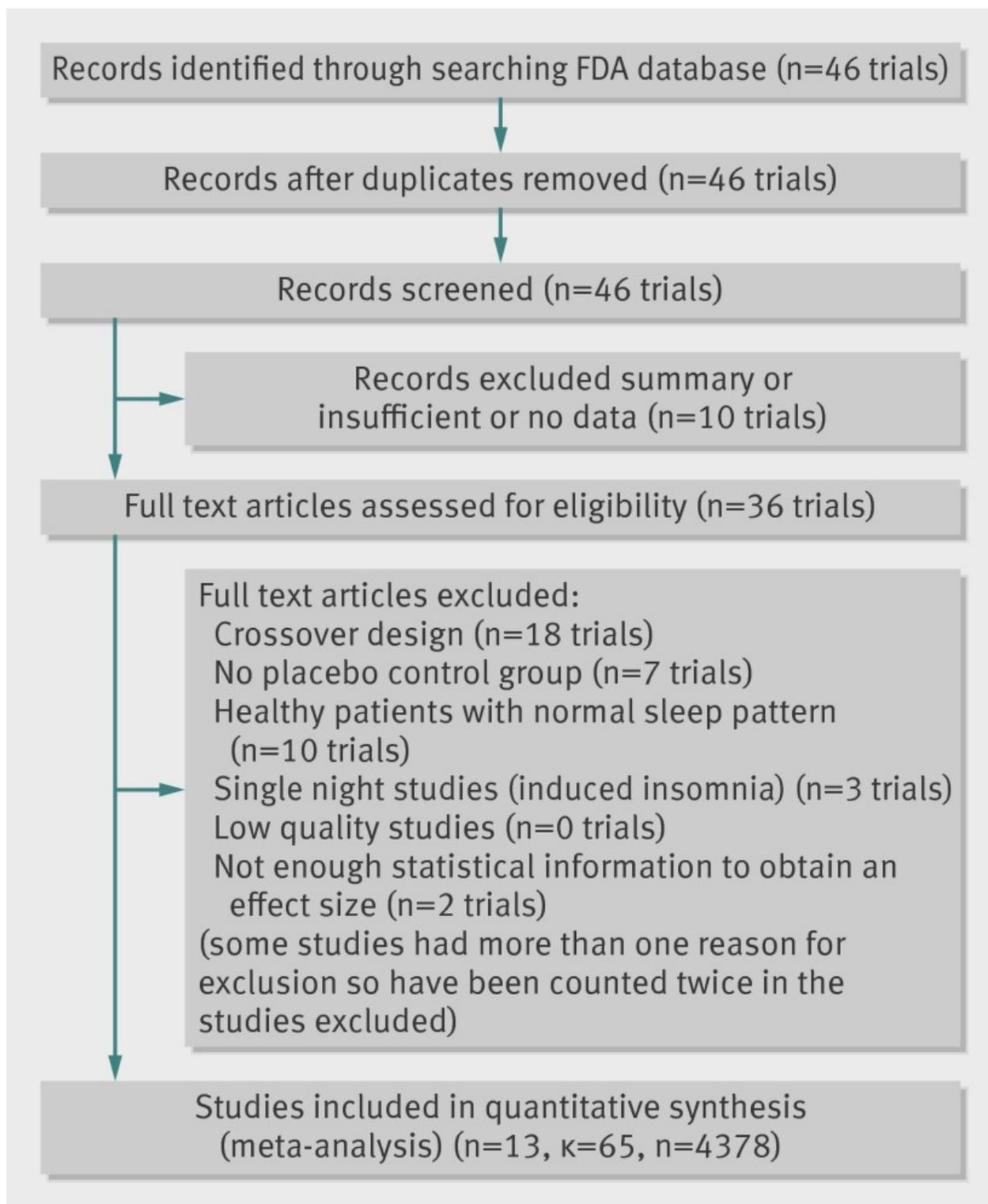


Fig 1 Identification of studies from FDA databases and inclusion in study

**Table 1**

Characteristics of studies included in review of Z drugs (a more detailed version of this table is in appendix 3)

Study identifier, year, country (No of sites)	Mean (SD)		age (years)	Drug	Recruitment	Baseline PSG (subjective)		Patient type	Outcome
	No	Women				Intervention:	Design		
EFC4529, 2004, US (29), Canada (5), Australia (6)	212	58%	44 (13)	Zolpidem MR 12.5 mg	Community	Intervention: 41.7 (61.4); placebo: 43.8 (62.0)	Phase II multicentre randomised double blind placebo controlled parallel group	Outpatients	Wake after sleep onset PSG, sleep latency PSG, No of awakenings PSG, No of awakenings subjective, sleep latency subjective, total sleep time PSG
EFC4530, 2004, Argentina (5), Canada (7), France (4), Germany (6), Mexico (2), US (16)	205	57%	70 (5)	Zolpidem MR 6.25 mg	Community	Intervention: 36.9 (56.0); placebo: 35.7 (62.9)	Randomised multicentre double blind placebo controlled	Outpatients	Wake after sleep onset PSG, sleep latency PSG, No of awakenings PSG, No of awakenings subjective, sleep latency subjective, total sleep time PSG
LSH17, 1988, US (4)	75	64%	38	Zolpidem MR 10 mg and 15 mg	Community	Intervention: zolpidem 10 mg: 35.8 (38.4); zolpidem 15 mg: 47.0 (61.0); placebo: 49.9 (70.4)	Double blind parallel group	Outpatients	Sleep latency PSG, Sleep efficiency PSG, No of awakenings PSG, sleep latency subjective, total sleep time subjective, No of awakenings subjective, sleep quality subjective.
LSH, 1992, US (6)	145	56%	45	Zolpidem (10 mg and 15 mg)	Community	Intervention: zolpidem 10 mg (65.1); zolpidem 15 mg (75.9); placebo: (58.2)	Double blind parallel group	Outpatients	Sleep latency subjective, total sleep time subjective, No awakenings subjective, sleep quality subjective
IV LSH, NR, NR	75	NR	NR	Zolpidem 10 mg and 15 mg	Community	Intervention: zolpidem 10 mg: 35.8 (57.0); zolpidem 15 mg: 47.0 (61.0); placebo: 49.9 (70.4)	Multicentre double blind randomised placebo controlled parallel group trial	Outpatients	Sleep latency PSG, sleep latency subjective, Sleep efficiency PSG



Study identifier, year, country (No of sites)	No	Women	Mean (SD) age (years)	Drug	Recruitment	Baseline PSG (subjective) sleep latency	Design	Patient type	Outcome
190-048, 2003, US and Canada (32)	234	57.7%	72.3	Eszopiclone 1 mg and 2 mg	NR	Intervention: NR; control: NR	Multicentre randomised trial	NR	wake after sleep onset PSG Sleep latency subjective, total sleep time subjective

NR=not reported; MR=modified release; PSG=polysomnographic.

\*Trial 307-1998 had two intervention arms: (i) zaleplon 10 mg for 14 days with outcomes measured at 7 days and 14 days compared with placebo and (ii) zaleplon 10 mg for 7 days followed by 20 mg for 7 days with outcomes measured at 7 days and 14 days compared with placebo; in both studies we used last measurement at 14 days and averaged dose at 15 mg as best approximation for study arm using 10 mg followed by 20 mg zaleplon.

PMW

PMW

PMW

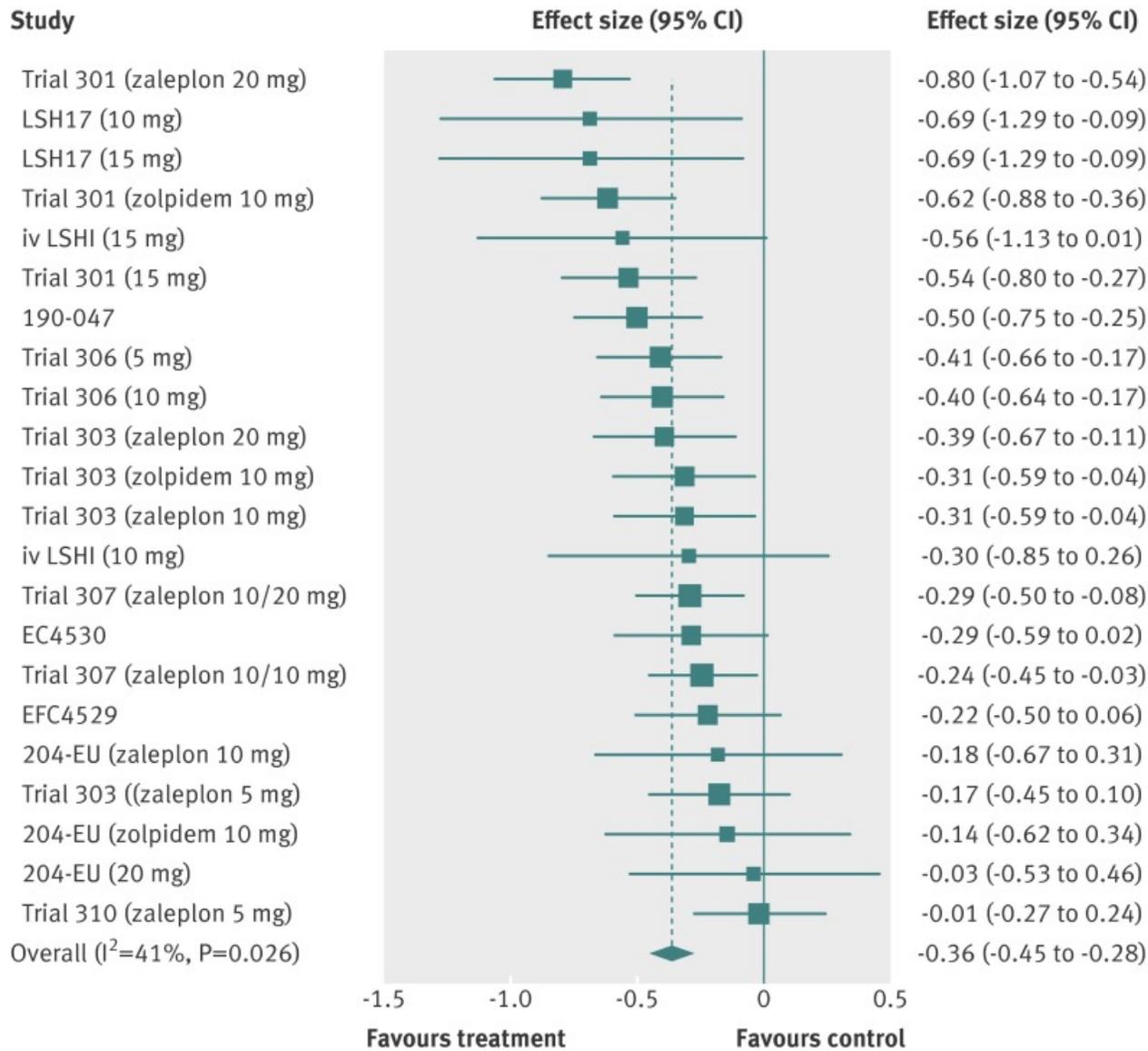


Fig 2 Forest plot for polysomnographic sleep latency under random effects assumptions

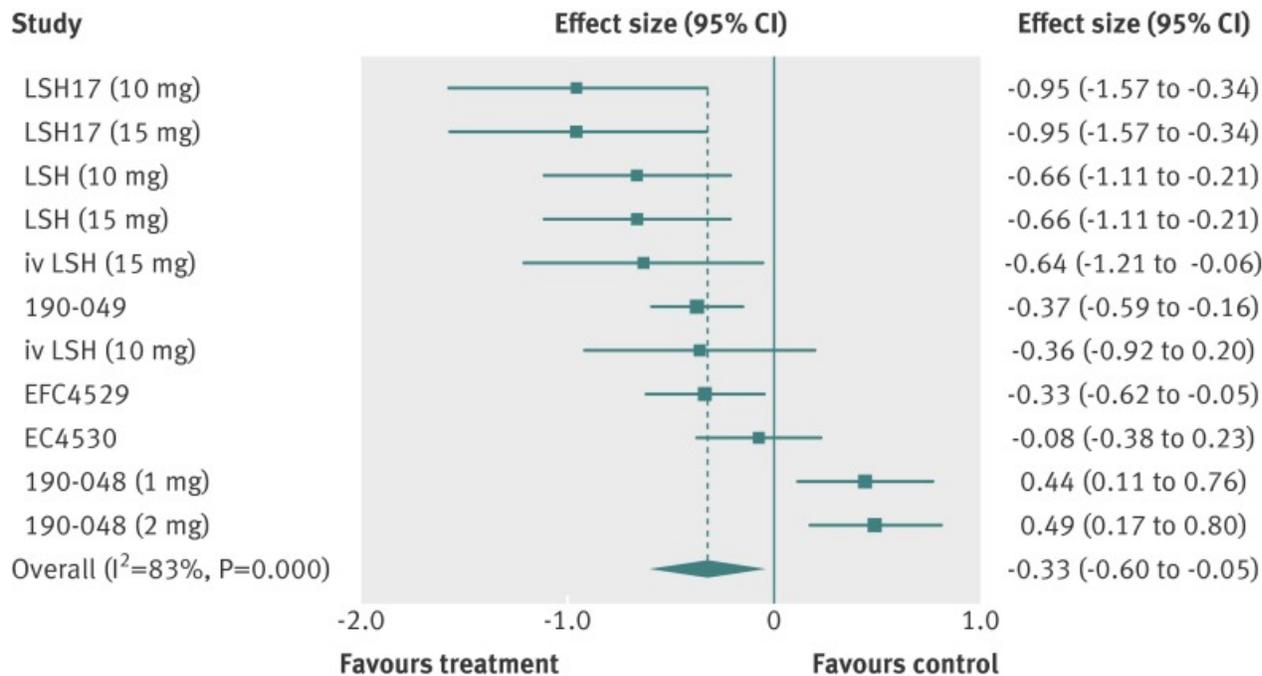


Fig 3 Forest plot for subjective sleep latency under random effects assumptions

**Table 2**

Weighted standardised mean differences in effect of Z drugs (treatment) or placebo on insomnia

	Weighted mean effect size (95% CI)							
	Within groups			Between groups		Homogeneity of effect sizes $I^2$ (95% CI)		
	No*	Treatment	Control	No*	Treatment v control	Treatment	Control	Treatment v control
<b>Primary outcome-sleep latency</b>								
PSG	16	-0.93 (-1.32 to -0.54)	-0.39 (-0.54 to -0.23)	22	-0.36 (-0.57 to -0.16)	89 (83.58 to 92.47)	0 (0 to 77.60)	41 (1.18 to 64.30)
Subjective	4	-0.67 (-1.30 to -0.03)	-0.33 (-0.63 to -0.03)	11	-0.33 (-0.62 to -0.04)	0 (0 to 49.26)	0 (0 to 66.15)	83 (70.98 to 90.05)
<b>Secondary outcomes</b>								
Wake after sleep onset (PSG)	2	-0.52 (-1.40 to 0.36)	-0.29 (-0.67 to -0.08)	3	-0.24 (-0.72 to 0.24)	83 (29.36 to 95.94)	50 (0 to 87.35)	0 (0 to 93.84)
No of awakenings (PSG)	2	-0.36 (-1.28 to 0.56)	-0.21 (-0.60 to 0.17)	4	-0.33 (-0.80 to 0.14)	91 (70 to 97.57)	0 (0 to 99.67)	65 (0 to 88.10)
No of awakenings (subjective)	2	-0.91 (-1.90 to 0.09)	-0.28 (-0.66 to 0.10)	6	-0.06 (-0.42 to 0.29)	87 (47.31 to 96.63)	36 (0 to 79.25)	85 (69 to 92.67)
Total sleep time (PSG)	2	1.06 (-1.37 to 3.49)	0.65 (-0.67 to 1.98)	2	0.41 (-0.51 to 1.32)	74 (0 to 94.19)	42 (0 to 83.14)	0
Sleep efficiency (PSG)	2	0.52 (-1.23 to 2.28)	0 (-0.59 to 0.59)	5	0.59 (-0.12 to 1.29)	54 (0 to 88.82)	0	0 (0 to 75.02)
Total sleep time (subjective)	0	—	—	7	0.45 (-0.08 to 0.98)	—	—	0 (0 to 71.12)
Sleep quality (subjective)	0	—	—	4	0.30 (-0.32 to 0.92)	—	—	0 (0 to 71.13)
Wake after sleep onset (subjective)	0	—	—	1	-0.16 (-0.60 to 0.28)	—	—	—

PSG=polysomnographic.

\*No of comparisons.

**Table 3**

Weighted mean raw differences in effect of Z drugs (treatment) or placebo on insomnia

	Weighted mean differences (95% CI)							
	Within groups			Between groups		Homogeneity of effect sizes $I^2$ (95% CI)		
	No*	Treatment	Control	No*	Treatment v control	Treatment	Control	Treatment v Control
<b>Primary outcome-sleep latency</b>								
PSG	14	-42 (-60 to -23)	-20 (-28 to -11)	14	-22 (-33.00 to -11.00)	96 (94.75 to 97.15)	41 (0 to 68.68)	94 (91.66 to 95.83)
Subjective	2	-24.99 (-30.06 to -19.92)	-19.43 (-26.61 to -12.25)	2	-6.90 (-26.00 to 12.37)	0 (0 to 100)	0 (0 to 100)	27 (0 to 72.41)
<b>Secondary outcomes</b>								
Wake after sleep onset (PSG)	2	-20 (-59 to 18)	-13 (-34 to 7.89)	2	-7.14 (-33.00 to 18.23)	65 (0 to 91.96)	63 (0 to 91.62)	0 (0 to 99.98)
No of awakenings (PSG)	2	1.24 (-6.34 to 3.89)	-0.94 (-12 to 9.99)	2	-0.47 (-5.12 to 4.17)	94 (81.24 to 98.13)	0 (0 to 99.98)	0 (0 to 99.90)
No awakenings (subjective)	2	2.88 (-7.15 to 1.39)	-1.05 (-4.86 to 2.76)	2	-1.77 (-4.61 to 1.07)	0 (0 to 100)	0 (0 to 99.95)	0 (0 to 99.81)
Total sleep time (PSG)	2	49.15 (-60 to 16)	35.10 (-34 to 10)	2	14.05 (-31.00 to 58.72)	63 (0 to 91.45)	61 (0 to 91.07)	0 (0 to 99.68)
Sleep efficiency (PSG)	1	4.27 (2.01 to 6.52)	0 (-2.52 to 2.52)	1	4.47 (2.08 to 6.86)	—	—	—
Total sleep time (subjective)	0	—	—	0	—	—	—	—
Sleep quality (subjective)	0	—	—	0	—	—	—	—
Wake after sleep onset (subjective)	0	—	—	0	—	—	—	—

PSG=polysomnographic.

\*No of comparisons.

**Table 4**

Moderator analysis for effect of Z drugs (treatment) or placebo on insomnia. Effect sizes are weighted standardised mean differences

Moderator variable†	Effect size (95% CI)	β‡
<b>Polysomnographic sleep latency</b>		
Dose (22 comparisons):		
1 mg	-0.24(-0.38 to -0.11)	-0.22*
20 mg	-0.50 (-0.64 to -0.35)	
<b>Subjective sleep latency</b>		
Year of data collection (9 comparisons):		
1988	-0.88 (-1.19 to -0.58)	0.63***
2004	-0.03 (-0.16 to 0.10)	
Age (9 comparisons) :		
38 years	-0.65 (-0.82 to -0.48)	0.89***
72 years	0.31 (0.13 to 0.50)	
Percentage of women (9 comparisons):		
55.9 %	0.01 (-0.17 to 0.18)	-0.40**
67.5 %	-0.67 (-0.99 to -0.34)	
Type of drug (11 comparisons):		
Eszopiclone (3 comparisons)	0.02 (-0.13 to 0.18)	-0.57**
Zolpidem (8 comparisons)	-0.47 (-0.61 to -0.33)	
Dose (11 comparisons):		
1 mg	0.13 (-0.37 to 0.30)	-0.70***
20 mg	-1.01 (-1.31 to -0.70)	

\*P<0.05; \*\*P<0.01; \*\*\*P<0.001.

†Effect size of each outcome was entered as dependent variable into separate weighted least squares regressions under fixed effects assumptions for each moderator variable independently; negative *ds* imply lower outcome at final available measures; estimates appear for observed extremes of continuous features.

‡Standardised regression coefficient.

Articles from The BMJ are provided here courtesy of **BMJ Group**