Comparison of Meta-Analytic Results of Indirect, Direct, and Combined Comparisons of Drugs for Chronic Insomnia in Adults: A Case Study

Ben W. Vandermeer, BSc, MSc, Nina Buscemi, PhD, Yuanyuan Liang, BSc, MSc, PhD and Manisha Witmans, Dip. ABSM, MD, FRCPC

Background: Our Center recently conducted a systematic review of the manifestations and management of chronic insomnia in adults. The efficacy and safety of benzodiazepines and nonbenzodiazepines, relative to placebo, were compared indirectly.

Objectives: Determine how the results of indirect comparisons made in the review compare with the results of direct comparisons, as well as with estimates derived from Bayesian mixed treatment comparisons. Establish general appropriateness of the use of results of indirect or mixed treatment comparisons.

Methods: Treatments were compared using frequentist direct, indirect, and combined methods, as well as Bayesian direct and mixed methods.

Results: Estimates for comparisons tended to be clinically and statistically similar across methods. Estimates obtained through indirect comparisons were not biased and were similar to those obtained through direct analysis.

Conclusions: Results of indirect comparisons made in the review, accurately reflected the current evidence. Frequentist and Bayesian methods of analysis of indirect comparisons should be considered when performing meta-analyses.

Key Words: meta-analysis, indirect comparison, mixed-treatment comparison

(*Med Care* 2007;45: S166–S172)

In July 2004, the University of Alberta/Capital Health Evidence Based Practice Center was commissioned by the Agency for Healthcare Research and Quality (AHRQ) and the Office of Medical Applications of Research (OMAR) to perform a systematic review of the efficacy and safety of treatments for chronic insomnia in adults.[1] We used meta-analysis to examine treatments such as medications [ie, benzodiazepines (BNZ) and nonbenzodiazepines (NBNZ)], behavioral therapy, complementary and alternative therapy (eg, melatonin, valerian), antidepressants, barbiturates, and alcohol. Through the analysis of randomized controlled trials (RCTs) that compared these treatments with placebo, we assessed the following outcomes: sleep onset latency (SOL), wakefulness after sleep onset (WASO), sleep efficiency (SE), total sleep time (TST), sleep quality (SQ), and adverse events (AE).

A stringent 11-month timeline for this systematic review forced the authors to make decisions about the main comparisons to be meta-analyzed with respect to the efficacy and safety of treatments for chronic insomnia in adults. We primarily compared active treatments with placebo, excluding from our review any RCT that did not include a placebo arm. To examine the relative efficacy and safety of active treatments, we performed indirect comparisons (eg, active treatment A vs. placebo compared with active treatment B vs. placebo) using the methods described by Bucher et al.[2] These results were treated as secondary outcomes.

In this method, a point estimate of the difference between 2 interventions is calculated by taking the difference of each intervention compared with a third intervention. Therefore, if T_{BP} is the direct estimate of a comparison between BNZ and placebo, and T_{NP} is the direct estimate of a comparison between NBNZ and placebo, then T_{BN} is the indirect estimate of a comparison between BNZ and NBNZ. Hence,

$$T_{BN} = T_{BP} - T_{NP}$$

$$VAR_{BN} = VAR_{BP} + VAR_{NP}$$

where VAR is the variance of the respective parameters. Because T_{BP} and T_{NP} are estimated from different studies, they are statistically independent.

This method of analyzing indirect comparisons is simple and straightforward and preserves the randomization of the initial RCTs. Although some have suggested that the method can lead to bias,[2] Song et al found that it generally does not.[3]

Though we included these results in our systematic review, we encouraged readers to treat them with more caution than they would the results of direct comparisons.[1] In the current study, we used both frequentist and Bayesian methods (see below) to compare the result of indirect com-

From the Department of Pediatrics, University of Alberta, Edmonton, Alberta, Canada.

Reprints: Benjamin Vandermeer, BSc, MSc, Department of Pediatrics, University of Alberta, Aberhart Centre One, Room 8412, 11402 University Avenue, Edmonton, Alberta, Canada. T6G 2J3. E-mail: bv1@ualberta.ca.

ISSN: 0025-7079/07/4500-0166

parison of BNZ and NBNZ with the results of direct and combined comparisons.

METHODS

Frequentist Estimates

We conducted all frequentist meta-analyses using the standard techniques of random effects meta-analysis of continuous variables using inverse variance as described by Deeks et al.[4] We extracted data from all RCTs with BNZ and NBNZ arms that had been excluded from the original meta-analysis because they lacked placebo arms. We combined these studies and the 3-arm trials (ie, trials analyzing BNZ, NBNZ, and placebo) included in the original review to calculate a direct estimate for the efficacy of BNZ versus NBNZ for the 6 outcomes of interest.

For each outcome, we also calculated an indirect estimate different from the estimate presented in the original review. In the calculation in the original review, the indirect estimate included all evidence from both 2- and 3-arm trials. In the current study, the indirect estimate included only 2-arm trials (ie, trials analyzing either BNZ vs. placebo or NBNZ vs. placebo). We excluded 3-arm trials because they were included in the direct comparison of BNZ versus NBNZ and we wished to avoid double-counting their data when we calculated the combined estimate.

We combined the direct estimate with the indirect estimate from the current study using the method described by Song et al.[3] The method weights by the inverse variance as in a standard meta-analysis.

For the mathematical details of these models, see Appendix A.

Bayesian Estimates

The use of Bayesian statistical methods in meta-analysis has been well documented.[5] These methods involve choosing prior distributions for the mean of the overall estimate, the means of the individual estimates of each study, and the between-study variance (when using random effects meta-analysis). The study data are then combined with the priors to derive posterior distributions for each study and for the overall estimated effect. We recalculated the direct comparison frequentist meta-analyses described above using these Bayesian methods.

The frequentist combined estimate of results of direct and indirect comparisons can be duplicated using Bayesian methods by a process known as the Mixed Treatment Comparison, described by Lu and Ades.[6] Under well-defined Bayesian hierarchical models proposed by Smith et al,[7] Bayesian approaches can be used to analyze simultaneously any number of interventions from a connected network of studies. One treatment is defined as a reference, and basic parameters are established by comparing other treatments to it. All other contrasts are then defined as functional parameters of these reference parameters.

The Bayesian models used for both direct meta-analysis and mixed treatment comparisons are outlined in detail in Appendix A. Although the references outline a binary model, the methods are easily adapted to continuous variables, because the binary data is converted into a log odds ratio and meta-analyzed as a continuous variable.

We derived posterior estimates for both the Bayesian meta-analysis and mixed treatment comparison using Gibbs Sampling via Markov Chain Monte Carlo simulation[7] in WinBUGS (Version 1.4). This method accounts for the correlation structure induced by 3-arm trials (or trials with more than 3 arms) and permits the user to estimate the probability that each intervention is the best of those chosen.

We gave all means vague prior distributions (ie, normal distribution with mean 0 and a sufficiently large variance). Choice of prior distributions for the between-studies variance parameter is more important than priors on the means, because posterior distributions are more sensitive to variance priors. An inverse gamma prior, which is weakly informative, was used as recommended by Gilks et al,[8] although research into which priors should be used for these parameters has been conducted and the results were largely inconclusive.[9] A sensitivity analysis was conducted to ascertain the effects the choice of prior had on the estimates.

RESULTS

Results of frequentist and Bayesian comparisons between BNZ and NBNZ for all 6 outcomes are presented (Table 1). One observation that is apparent in nearly all 6 outcomes is that there are more studies involved in the indirect comparisons than there are in the direct comparisons, yet the confidence intervals in the latter group are not much narrower than the former—in fact in some instances they are wider. This finding can by explained as the following. It can be shown mathematically that assuming all study standard errors and between-study variances are approximately equal (σ), a direct comparison involving n studies would have standard error proportional to $\sigma/\sqrt{n}$, whereas an indirect comparison involving n studies would have standard error $2\sigma/\sqrt{n}$. Thus it takes on average 4 times as many studies in an indirect comparison to match the variance of a similar direct comparison (ie, $2\sigma/\sqrt{4n} = \sigma/\sqrt{n}$) . The increase in confidence interval width that can be observed between the modified and original indirect comparisons is attributable to both the decreased sample size as well as the elimination of 3-arm trials that tended to have lower variation.

Sleep Onset Latency

Results were clinically and statistically similar across comparisons. There was less than 1 minute difference in sleep onset across the 6 comparisons, and an insignificant difference between BNZ and NBNZ for each comparison. The Bayesian rankings confirmed the uncertainty of the relative efficacy of the 2 active interventions, indicating a 75% likelihood that NBNZ was more efficacious and a 25% likelihood that BNZ was more efficacious. The rankings showed that each treatment was superior to placebo.

Wakefulness After Sleep Onset

WASO values among the 6 comparisons ranged from a 10-minute advantage for BNZ (observed with the frequentist indirect approach) to a 2-minute advantage for NBNZ (with

TABLE 1. Results of Comparisons Between Benzodiazepines and Nonbenzodiazepines for 6 Outcomes

	Frequentist				Bayesian		
Outcome	**Direct Comparison**	**Indirect Comparison**	**Modified Indirect Comparison**	**Combined**	**Direct Comparison**	**Combined**	**Probability of "Best"**
Sleep onset latency (min)	1.58 (−3.38, 6.54) 11 Studies	1.63 (−4.35, 7.61) 62 Studies	2.20 (−4.64, 9.04) 54 Studies	1.79 (−2.22, 5.81) 65 Studies	1.62 (−4.81, 6.88) 11 Studies	1.74 (−3.33, 6.62) 65 Studies	NBNZ: 75.5%, BNZ: 24.5%, Placebo: 0.0%
Wakefulness after sleep onset (min)	−0.46 (−19.88, 18.97) 3 Studies	−10.47 (−26.78, 5.84) 17 Studies	−8.27 (−25.57, 9.03 16 Studies	−4.82 (−17.74, 8.10) 19 Studies	1.99 (−28.02, 25.24) 3 Studies	−3.91 (−19.80, 9.68) 19 Studies	NBNZ: 27.3%, BNZ: 72.7%, Placebo: 0.0%
Sleep efficiency (% points)	3.39 (1.29, 5.49) 3 Studies	0.48 (−2.24, 3.20) 16 Studies	−1.60 (−5.56, 2.36) 13 Studies	2.29 (0.44, 4.15) 16 Studies	3.40 (0.39, 6.42) 3 Studies	1.50 (−1.12, 4.09) 16 Studies	NBNZ: 12.4%, BNZ: 87.6%, Placebo: 0.0%
Total sleep time (min)	12.38 (0.64, 24.12) 8 Studies	11.13 (−2.51, 24.77) 37 Studies	6.66 (−10.27, 23.59) 31 Studies	10.52 (0.88, 20.17) 39 Studies	10.53 (−0.38, 24.50) 8 Studies	9.98 (−0.53, 20.26) 39 Studies	NBNZ: 96.9%, BNZ: 3.1%, Placebo: 0.0%
Sleep quality (SMD)	0.11 (−0.04, 0.27) 11 Studies	0.32 (0.14, 0.50) 45 Studies	0.35 (0.13, 0.57) 38 Studies	0.19 (0.06, 0.32) 49 Studies	0.11 (−0.06, 0.28) 11 Studies	0.22 (0.08, 0.37) 49 Studies	NBNZ: 99.9%, BNZ: 0.1%, Placebo: 0.0%
Adverse events (risk difference)	0.07 (0.00, 0.14) 17 Studies	0.10 (0.04, 0.16) 59 Studies	0.11 (0.03, 0.19) 52 Studies	0.09 (0.03, 0.14) 69 Studies	0.07 (−0.03, 0.18) 17 Studies	0.09 (0.03, 0.16) 69 Studies	NBNZ: 1.7%, BNZ: 0.0%, Placebo: 98.3%

Direct comparison includes results of meta-analyses using only studies that compared the 2 treatments directly. Indirect comparison includes results as calculated in the original systematic review. All studies that contained 1 (or both) of the 2 interventions in addition to a placebo arm are included. Modified indirect comparison is same as indirect comparison, but with the 3-arm trials removed from the analysis. Combined estimate (frequentist): weighted combination of the direct comparison and the direct and modified indirect comparisons. Combined estimate (Bayesian): Bayesian mixed treatment comparison of all studies. Probability of "Best": The percentage that each intervention ranked number 1 among the 3 in terms of efficacy or safety in each iteration of the Monte–Carlo Markov-Chain simulation.

NBNZ indicates nonbenzodiazepines; BNZ, benzodiazepines; SMD, standardized mean difference.

the Bayesian direct approach), although none of these differences was statistically significant. The difference in the estimates for direct and indirect analyses of WASO using the frequentist approach (less than a 1-minute advantage for BNZ was observed with the frequentist direct approach, whereas a 10-minute advantage was observed with the frequentist indirect approach) may be clinically important in the context of other clinical outcomes. For example, a 10-minute increase in WASO with a greater number of awakenings may be clinically important, whereas a 10-minute increase in WASO without an increase in the number of awakenings may not be clinically important. The Bayesian rankings demonstrated a 73% likelihood that BNZ was more efficacious and a 27% likelihood that NBNZ was more efficacious.

Sleep Efficiency

The results for SE (defined as the percentage of total time in bed that a subject was asleep) ranged from an advantage of 3.4 percentage points for BNZ (observed with the direct comparison for both methods) to an advantage of 1.6 percentage points for NBNZ (observed with the frequentist modified indirect comparison). The differences among methods were not clinically significant. Also, the direct estimates indicated a statistically significant difference between the 2 interventions, unlike the indirect estimates, which indicated an insignificant difference. The Bayesian rankings suggested an 88% likelihood that BNZ was the most efficacious of the interventions.

Total Sleep Time

Total sleep time did not differ substantially across comparisons. Some estimates were statistically significant and some were statistically insignificant, but all were clinically comparable. The Bayesian rankings suggested a 97% likelihood that NBNZ was the most efficacious intervention.

Sleep Quality

Differences in SQ (defined as subject's overall satisfaction with sleep) estimates across the 6 comparisons were statistically insignificant. Clinical significance was more difficult to judge because the standardized difference method required data to be represented in units of standard deviation. We also note that the indirect evidence showed a statistically significant difference in SQ between interventions, whereas the direct evidence differences were not statistically significant. Despite the insignificance of some of the other estimates, the Bayesian rankings indicated almost a 100% likelihood that NBNZ was the most efficacious intervention.

Adverse Events

The AE estimates were similar across comparisons, ranging from risk differences of 0.07 (observed with the direct comparison with both methods) to 0.11 (observed with the frequentist modified indirect comparison). None of the differences were statistically significant. The clinical importance of the difference in AE depends on the nature and severity of the AEs. Because most AEs in this study were mild, the differences were not clinically significant. Only 1 comparison (Bayesian direct) indicated a statistically insignificant difference between BNZ and NBNZ. The Bayesian rankings demonstrated that placebo resulted in the fewest AEs.

Prior Sensitivity for Bayesian Analyses

All our Bayesian analyses were completely insensitive to the vague priors chosen for the basic trial level and combined means for each group. We used normal priors centered at 0 with large variance, and our results did not change when we modified the variance or used uniform priors set to similar limits.

Results were much more sensitive to the choice of prior distribution on the between-study variance parameter τ^2. Lambert et al[9] suggest several priors that can be used. As mentioned in Appendix A, we chose the inverse gamma for our primary analysis but we also analyzed a uniform distribution on the standard deviation. Weakly informative versions of both priors were also tested to check for sensitivities to selection of the parameters.

Not surprisingly, between-study variance prior sensitivity was directly related to the number of studies in the analysis, especially for the direct meta-analyses. For SOL, SQ, and AE the effect was very minimal in both direct and combined meta-analyses, mainly due to the large amount of studies involved in these comparisons. TST, WASO, and SE also contained negligible differences for the combined analysis, but some sensitivity was noticed in the direct meta-analyses, particularly in WASO and SE which had only 3 studies included. Reducing the variability of the gamma distribution (going from $\alpha,\beta = 0.001$ to $\alpha,\beta = 0.1$) had little to no effect, but using a vague uniform prior [U(0,100)] did somewhat increase the credible interval for these 3 outcomes. When the uniform prior was made weakly informative [U(0,10) and U(0,5)], the estimates and credible intervals were more in line with the gamma priors. The low amount of studies in these outcomes makes it very difficult to estimate between-study variance from the data, thus the priors become influential in the final results. The U(0,100) prior allows unreasonably high values to be sampled for the variance and should probably not be used when sample sizes are small (under 10 studies). See Lambert et al[9] for more details.

Comparison of Direct and Indirect Evidence

Figure 1 displays the most important direct and indirect comparisons: frequentist direct versus frequentist indirect estimates, frequentist direct versus frequentist combined estimates, and Bayesian direct versus Bayesian mixed treatment comparison estimates. None of the results for any comparison were statistically significant. Furthermore, the differences across methods in SOL, WASO, TST, and AE were small relative to their effect sizes. A modest difference across methods was observed in SE and SQ.

DISCUSSION

Clinical Interpretation of the Findings

Judging the clinical importance of differences in pooled estimates derived by indirect, direct, and combined analyses is difficult because, for sleep parameters such as SOL and WASO, no clinically important effect threshold has been

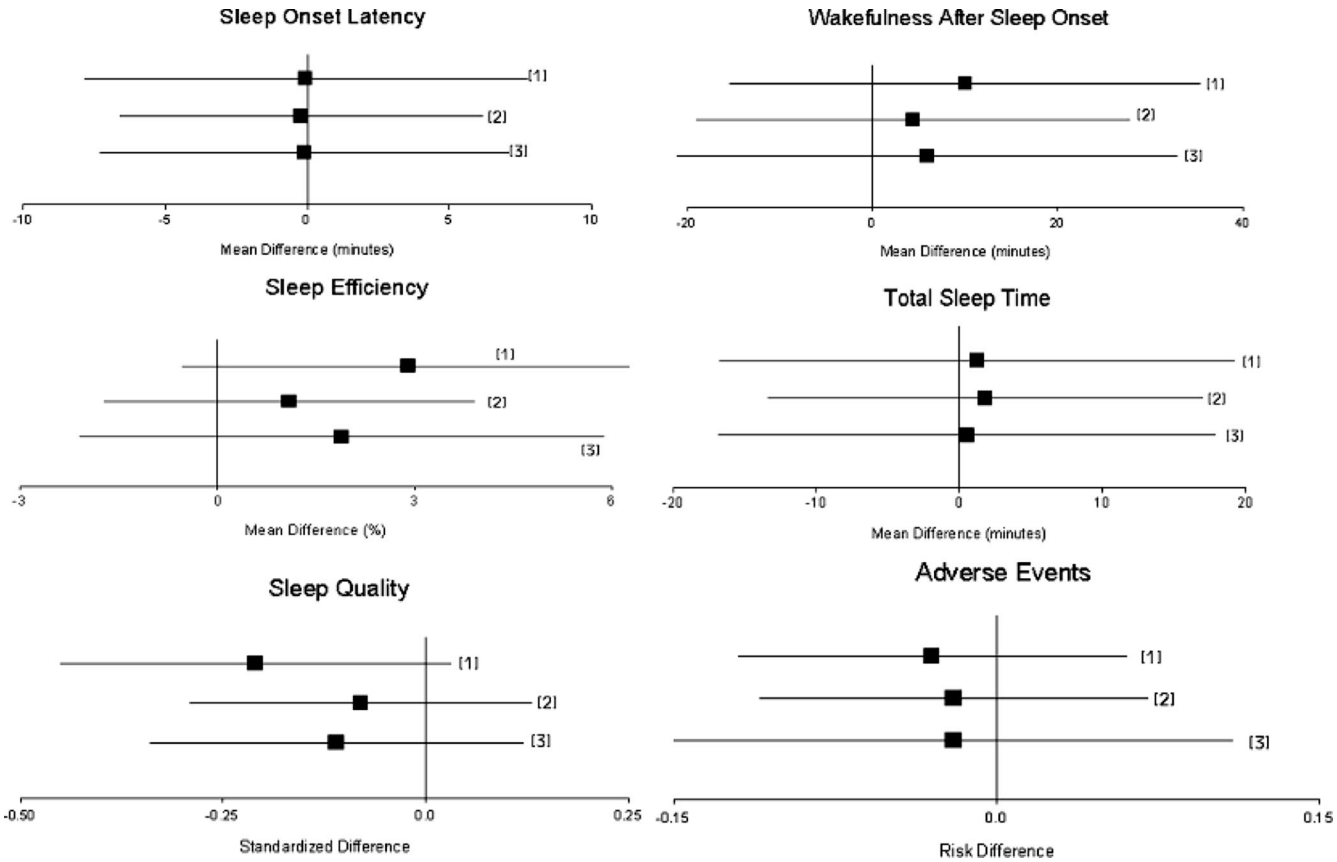

FIGURE 1. Comparison of results of direct and indirect comparisons for 6 outcomes. (1) Comparison of direct frequentist method versus indirect frequentist method. (2) Comparison of direct frequentist method versus combined frequentist method. (3) Comparison of direct Bayesian method versus combined mixed treatment Bayesian method.

established. Furthermore, SE, measured as time asleep relative to time in bed, is a crude measure of adequate or restorative sleep. Clinical interpretation of these findings is based on both individual judgment and consideration of related factors, such as the severity of the disorder at baseline, the patient's perception of improvement, and the number of awakenings.

In our clinical opinion, the differences among estimates for SOL, SE, and TST are not clinically important because of their small magnitude. The differences in numbers of AEs are also clinically unimportant because the AEs were generally mild (eg, headache, nausea). The differences in WASO estimates could be considered clinically important in the context of other clinical outcomes, such as number of night awakenings, but they are not significant when considered in isolation. The differences in SQ estimates are more difficult to assess clinically because they are represented in units of standard deviation and thus must be judged from a statistical rather than a clinical perspective.

Statistical Interpretation of the Findings

We found no significant differences in any of the comparisons made among direct, indirect, and combined evidence. Nonetheless, some minor clinical differences emerged (in terms of point estimate magnitude) between estimates, as noted above. Changes in statistical conclusions also occurred. In some cases, a direct comparison yielded an insignificant result while an indirect comparison yielded a significant result, or vice versa. The former situation is discussed above; the latter can be explained with differences in confidence interval sizes, because shifts in the point estimates were minor. Changes in confidence intervals can be attributed to differences in the number of studies used in the analysis for each estimate and to the wider confidence intervals of indirect estimates than of direct estimates (assuming similar sample sizes). In addition, the Bayesian estimates tended to have wider confidence intervals because the Bayesian random effects inferences used in this study use a varying between-studies variance parameter, unlike the standard DerSimonian and Laird frequentist estimate, which assumes a constant known between-study variance.

CONCLUSIONS

In terms of the insomnia systematic review, we believe that the indirect comparisons provided were reasonable approximations of the direct comparisons that we could have performed. Perhaps we need not have been so modest in our presentation of these earlier indirect estimates. Although they had been calculated for all outcomes, only results for SOL and AEs were given in the text.[1] Indirect evidence could have

been used in the absence of (or combined with) direct evidence and the results would not have differed much clinically or statistically from the direct results. Based on our 1 case study, it is difficult to make statements on how generalizable our results may be, but combined with Song's[3] results, there is certainly reason for optimism. At the very least, indirect evidence can be included as a sensitivity analysis in a standard frequentist meta-analysis.

When multiple interventions need to be analyzed, the Bayesian mixed treatment approach should be strongly considered as the method. In our example, results derived by Bayesian methods did not differ greatly from those derived by frequentist methods, but they can and did add value to the analysis, as demonstrated by the ranking statistic. Bayesian methods are also tidier when more than 3 interventions are examined. Using a frequentist method with 3 interventions is reasonably straightforward, but when a fourth intervention is introduced, combining direct and indirect evidence becomes messy. With mixed treatment comparisons, many interventions can be easily assimilated into the analysis. All pairwise contrasts can be isolated and other information can be obtained. The ranking of interventions becomes particularly useful in this instance, as it yields a straightforward method of simultaneously analyzing all interventions. By adding a cost function to the program, it is easy to incorporate a cost-effectiveness analysis as well. Other optional useful outputs include a point estimate and confidence interval for a "typical study," and an alternative estimate for each study that "borrows strength" from the other studies, and may be considered a better reflection of the study population.[10]

Those who conduct many meta-analyses may fall into a trap of using the same methods for each, but it is important to remember that each contains different populations, study designs, interventions, and outcomes that should be considered when choosing methods. Song et al[3] have shown that indirect evidence, properly analyzed, most often yields results that do not differ significantly from those obtained via equivalent direct evidence. This suggests there is no reason not to incorporate indirect evidence into an analysis. Using indirect evidence becomes particularly important when direct evidence is scarce or difficult to obtain and indirect evidence is abundant.

REFERENCES

1. Buscemi N, Vandermeer B, Friesen C, et al. Manifestations and management of chronic insomnia in adults. *Evid Rep Technol Assess (Summ)*. 2005;125:1–10.
2. Bucher HC, Guyatt GH, Griffith LE, et al. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol*. 1997;50:683–691.
3. Song F, Altman DG, Glenny AM, et al. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ*. 2003;326:472–476.
4. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Smith GD, Altman DG, eds. *Systematic Reviews in Health Care: Meta-Analysis in Context*. London, England: BMJ Publishing Group; 2001:285–312.
5. Sutton AJ, Abrams KR, Jones DR, et al. Bayesian methods in meta analysis. In: *Methods for Meta-Analysis n Medical Research.* Chichester, England: Wiley; 2000:163:–190.
6. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23:3105–3124.
7. Smith TC, Spiegelhalter DJ, Thomas SL. Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med*. 1995;14:2685–2699.
8. Gilks WR, Richardson S, Spiegelhalter DJ. *Markov Chain Monte Carlo in Practice*. London, England: Chapman & Hall; 1996.
9. Lambert PC, Sutton AJ, Burton PR, et al. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Stat Med*. 2005;24:2401–2428.
10. Higgins JP, Whitehead J. Borrowing strength from external trials in a meta-analysis. *Stat Med*. 1996;15:2733–2749.

APPENDIX A

Models and Computations Used for Meta-Analytic Estimates

Direct Comparison—Frequentist

From each study i we have μ_i and SE_i representing the mean difference between the 2 groups being evaluated with its respected standard error. The standard fixed effect, inverse variance combined meta-analytic estimate (μ_{IV}) is given by:

$$\mu_{IV} = \frac{\sum w_i \mu_i}{\sum w_i}$$

where $w_i = \frac{1}{SE_i^2}$

For the random effects estimate we calculate further parameters τ^2 and w_i' as:

$$\tau^2 = \frac{Q - (k-1)}{\sum w_i - \left(\frac{\sum w_i^2}{\sum w_i}\right)}$$

if $Q > (k-1)$ and 0 otherwise.

$$w_i' = \frac{1}{SE_i^2 + \tau^2}$$

where k is the number of studies and Q is the heterogeneity statistic: $Q = \sum w_i(\mu_i - \mu_{IV})^2$

The pooled random effects estimate and its standard error are then given by:

$$\mu_{RE} = \frac{\sum w_i' \mu_i}{\sum w_i'}$$

and

$$SE(\mu_{RE}) = \frac{1}{\sqrt{\sum w_i'}}$$

These are the values used for our frequentist direct comparison.

Indirect Comparison—Frequentist

Given any meta-analytic comparison of A versus B and another comparison of A versus C, one can obtain an indirect comparison of B versus C. Let μ_{AB} and SE(μ_{AB}) be the combined mean difference and corresponding standard error of the A versus B meta-analytic comparison and let μ_{AC} and SE(μ_{AC}) be the equivalent values for the A versus C comparison. The comparison of B versus C is estimated as:

$\mu_{BC} = \mu_{AB} - \mu_{AC}$ with corresponding standard error

$$SE_{BC} = \sqrt{SE_{AB}^2 + SE_{AC}^2}$$

Combined Comparison—Frequentist

Given an estimated meta-analytic mean difference and standard error with a corresponding indirect estimate with standard error, a combined estimate can be derived using the techniques of random effects meta-analysis. The direct estimate is considered one "study" whereas the indirect estimate is considered a second "study." The random effects procedure is then followed as a normal meta-analysis of 2 studies.

Direct Comparison—Bayesian

The model used for the Bayesian meta-analyses was as follows. Given mean differences d_i for each study with respected standard errors SE_i, this data is used as the likelihood function of standard Bayesian updating equation: $P(\theta \mid data) \propto P(\theta)P(data \mid \theta)$. The exact model was:

$d_i \sim Normal\ (\mu_i, SE_i^2)$ (likelihood function).

$\mu_i \sim Normal\ (d, \tau^2)$ (distribution of individual study means μ_i).

$d \sim Normal\ (0,10000)$ (prior distribution on overall mean d).

$\frac{1}{\tau^2} \sim Gamma\ (0.001,0.001)$ (prior distribution on between-study variance τ^2).

Thus we chose a normal distribution centered at 0 for the prior on location parameter d and an inverse gamma parameter for the between-study variance parameter τ^2. The parameters for the variance of d and the α and β parameters of the γ distribution were based upon,[9] and were examined in sensitivity analyses.

Mixed Treatment Comparison—Bayesian

This model combines data from studies that contain any number of interventions and gives estimates of all pairwise comparisons without distinguishing between direct and indirect evidence. Because we have more than 2 treatments, it is not possible to isolate 1 comparison (although for the purposes of this study we focused on the benzodiazepine versus nonbenzodiazepine comparison). We define d_{BP} as the comparison of benzodiazepine with placebo and d_{NP} as the comparison of nonbenzodiazepine with placebo. We define b_i—the baseline treatment of each study (1 = placebo, 2 = benzodiazepine, 3 = nonbenzodiazepine)—as the "lowest" treatment of each study. We get our mixed treatment model:

$\mu_{ik} \sim Normal\ (m_{ik}, SE_{ik}^2)$ Likelihood function (treatment k in study i)

$\delta_{ibk} \sim Normal\ (d_{bk}, \sigma^2)$ Main model for treatment effects.

$\mu_{ib} \sim Normal(0,10000)$ Prior distribution for study level estimates.

$d_{BP}, d_{NP} \sim Normal\ (0,10000)$ Prior distribution for 2 functional parameters.

$\frac{1}{\sigma^2} \sim Gamma\ (0.001,0.001)$ Prior distribution for between-study variation.

The chosen prior distributions were examined in sensitivity analyses.