Reviewer’s report

Title: Spontaneous improvement in randomised clinical trials: meta-analysis of three-armed trials comparing no-treatment, placebo and active intervention

Version: 2 Date: 4 October 2008

Reviewer: Jesse Berlin

Reviewer’s report:

Major Compulsory Revisions

1. This is essentially a restatement of my previous comment number 5, and is prompted, in part, by the strength of Dr. Lau’s concerns about what I think is the same issue. We all agree that the analyses you did, fail to take the randomization into account. I’ll say again that my suspicion is that the choice of method won’t make a huge difference, but I’m still left wondering if there isn’t some kind of sensitivity analysis that could be performed to assess the robustness of the results to the choice of statistical approach. The variability of the various effects across the clinical areas, coupled with the high I-squared values within some of the areas, still leaves me a bit uncomfortable without seeing some kind of confirmatory analysis.

I’ll suggest again, and remain open to arguments as to why this approach might be completely inappropriate, that it might be possible to get a within-study measure of the relative contributions. Essentially, I’m suggesting that you repeat the calculation you did of percent contribution, but do that within each trial. These are all 3-arm studies, so the effect of spontaneous improvement, within a study, could be defined as “change from baseline on no treatment / change from baseline on active treatment.” Similarly, you could calculate the placebo contribution within a study, as you did using the mean effects. As I noted in my previous comments, your results are based on the “ratio of the means” rather than the “mean of the ratios.”

There are reasons (e.g., the overall allocation ratios being about 1.0) to believe that this (or another related) analytic approach would yield similar results. There are also weaknesses with this proposed approach. One that comes to mind immediately is the problem of working on the ratio scale. One might reasonably argue that the “average” of a ratio of 2 and a ratio of 0.5 should be 1.0, which would come from averaging the log ratios, then exponentiating (i.e., taking the geometric mean), and not the simple arithmetic mean of 1.25. Working on the log scale would, admittedly, make the exercise far more complicated to interpret.

So again, I offer this suggestion as a potential approach to addressing the concerns about the use of separate treatment arms, ignoring stratification by study.
Minor Essential Revisions

2. It’s strictly a formatting point, but you should be consistent about the use of commas versus decimal points in the tables.

Discretionary Revisions

3. Just a small point about presentation. As an alternative to the current figure 1, I wonder if a bar chart might work. For the overall effect, you would have a single bar, with 3 sections. The bottom section would be 20%, representing the contribution of spontaneous improvement. The middle section would be 25%, representing the “corrected” contribution of placebo. These two sections total to the 45% (with rounding, this is the 47% you report). The top section would be the remaining 55%, representing the effect of active treatment. You might even consider presenting a separate bar for each of the clinical areas (a visual representation of the current table 2, which would nicely demonstrate the variability across those clinical questions.)

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I am a full-time employee of Johnson & Johnson Pharmaceutical Research and Development. I know of no conflicts directly related to this methodologic paper.