References


Reply

To the Editor:

We thank Dr. Friedman for his interest in our review of magnetic resonance imaging findings in schizophrenia. We also agree with his point that “reasonable people can differ about the appropriate application of meta-analytic techniques,” and he does not quibble about our not using meta-analysis. More specifically, we noted in our review (McCarley et al 1999, 1100) that for a review of this broad a scope, covering more than a decade and all published studies, meta-analysis was not the most appropriate approach, as it would yield a false sense of numerical exactness. We suggest that one cannot simultaneously do a comprehensive review of a decade of studies, the goal of this review, and also perform a valid meta-analysis, unless the variables [of technical quality, ROI definition, and subject moderator variables] have remained constant, a constancy for which there is little evidence [italics added].

Where Dr. Friedman does have a problem with our review is in his reading our review as endorsing a simple-minded presentation of vote counting; in his reading we “fail to consider the pros and cons.” We italicize his to emphasize that, in fact, the review does note the pros and cons of this approach. Moreover, we used a method that, had Dr. Friedman read the relevant parts of the review, would have obviated his concern of a less than 10^{-20} probability of a correct conclusion, given an effect size of .4 and 10^6 studies in his example.

We call the reader’s attention to the section of our review on page 1100, beginning with: “Second, we need to comment on the tabular presentation of the studies in Table 1.” (We go on to note the need to provide literature references to back up review “expert opinions” so that the reader can judge the evidence. We then take up the problems of using meta-analysis in a comprehensive review of this nature. We encourage the reader to reread this section, since space prevents our repeating the details of argument here, and since Friedman did not contest this point.)

The key points that Dr. Friedman apparently missed are on the same page:

We, nonetheless, are sympathetic to readers who would like some more quantitative information and for this reason we have provided a table of subject N for each study in a summary table. It will be clear that the statistical power of negative studies with small subject N is less, and that their results consequently are less convincing than those negative studies with a larger N. We also, following a suggestion of Rosenthal (1987), computed the probability of the observed positive and negative statistical findings, using a two-tailed probability, and the alpha level of p < .05 of the studies. The reader will note that this procedure, like the more standard meta-analyses, assumes comparability of the studies with respect to measurements and subjects, although it does not assume normality of the distributions. We use this probability statistic with caution, since, in our opinion, it is hazardous to assume that the studies are comparable in methodology and subject characteristics; however, it may be of interest to the reader that, if one does assume comparability, the binomial theorem computation (using p < .05 for a positive study) shows that all ROIs surveyed in Table 1 show a two-tailed p < .05 for the number of positive studies, except for the fourth ventricle and cerebellum, and all ROIs are less than or equal to .002 except for the occipital lobe (.004). . . .

Although this procedure has the same liabilities as meta-analysis (of which it is a nonparametric variant) in comparing possibly noncomparable studies, it is useful, we believe, because it gives the reader another point to figure into his/her weighting of the data and, of course, would avoid the conclusion Friedman introduces in his example. What he labels as a simple vote count was, in fact, not.

Let us go through the computations. The probability of a statistically significant result in a single study is p < .05. The probability of 10 studies each showing statistically significant results is [p < .05 × 10^{-20}]^{10} = p < .00000000000009765625, or 9.765625 × 10^{-14}.

The reader and Dr. Friedman will recall that, for the case of a series of mixed successes and failures, the binomial theorem provides the correct computation of the probability for p successes and q failures when the true probability of a success is <.05, and this is the way we calculated the probabilities cited above for the summary in Table 1.

Thus, using this formula (Z approximation, since N is large) for Dr. Friedman’s example, the calculated probability of concluding there was no significant effect is <10^{-14}, a probability level that would give the reader a clear indication of which way the literature was pointing. Dr. Friedman incorrectly calculated that the probability of correctly finding a significant effect was
about $10^{-20}$. Since the binomial calculation shows the probability of not finding the correct effect is $<10^{-6}$, Dr. Friedman falls into error by a very conservatively estimated factor of $10^{26}$.

Nonetheless, we are grateful to Dr. Friedman for raising the issue of how best to summarize the literature, an important issue for all fields, and one that is useful for the readers of Biological Psychiatry to consider.

Robert W. McCarley
Cynthia G. Wible
Melissa Frumin
Yoshio Hirayasu
James Levitt
Martha E. Shenton

Harvard Medical School
VA Boston Healthcare System (Brockton Division)
Department of Psychiatry (116A)
940 Belmont Street
Brockton MA 02301

References