

under selection, I do not believe that estimates from any one of these models should be taken too seriously. Estimates from a variety of different selection models are very valuable, however, as a way to assess the sensitivity to selection effects of conclusions derived from a body of research. It may be important to know, for example, that a realistic selection model would lead to a combined estimate of treatment effect that is only half as large as that observed in published studies. This can easily happen if most of the observed effects have p -values only slightly smaller than the critical p . It is also important to know that no reasonable selection model has much effect on the combined estimate of treatment effect. This can happen when most of the observed effects have very small p -values. By viewing selection models as techniques for sensitivity analysis, we may exploit them more effectively

in the attempt to draw scientific conclusions from collections of related research studies.

ADDITIONAL REFERENCES

- CHAMPNEY, T. F. (1983). Adjustments for selection: Publication bias in quantitative research synthesis. PhD dissertation, Univ. Chicago.
- HEDGES, L. V. (1984). Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *J. Ed. Statist.* **9** 61–85.
- LAIRD, N. M. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *J. Amer. Statist. Assoc.* **73** 805–811.
- LAIRD, N. M. (1982). Empirical Bayes estimates using the nonparametric maximum likelihood estimate for the prior. *J. Statist. Comput. Simulation* **15** 211–220.
- LAIRD, N. M and LOUIS, T. A. (1987). Empirical Bayes estimates based on bootstrap samples. *J. Amer. Statist. Assoc.* **82** 739–750.

Comment: Assumptions and Procedures in the File Drawer Problem

Robert Rosenthal and Donald B. Rubin

Interesting and important questions have been raised about the file drawer problem in the thoughtful and constructive contribution by Iyengar and Greenhouse. Our purpose here is to (a) examine the assumptions underlying the file drawer computations, (b) report some empirical estimates of retrieval bias relevant to these computations, (c) report the results of a study of retrieval bias in an early and fully documented meta-analysis and (d) comment on the framework described by Iyengar and Greenhouse and other frameworks relevant to meta-analysis.

1. ASSUMPTIONS UNDERLYING THE ORIGINAL FILE DRAWER COMPUTATIONS

Iyengar and Greenhouse stated that the file drawer computations (Rosenthal, 1979) are "...based upon the assumption that the unpublished studies are in fact a random sample of all studies that were done." This is, however, *not* the assumption underlying the file drawer computations proposed in Rosenthal (1979). Rather, Rosenthal (1979) explicitly assumed

that (a) the null hypothesis is true (expected mean $z = 0.00$) and (b) the selection process is such that all results significant at say, .05, *two-tailed*, are published (or retrieved) whereas those that are not significant are not published (or not retrieved).

In their own assumptions underlying the file drawer computations, Iyengar and Greenhouse assume the same null hypothesis but, when critically evaluating the file drawer computations, their selection process assumption is that all results significant at say, .05, *one-tailed*, are published (or retrieved), while those that are not significant in that direction are not published (or not retrieved). In their formal models, however, Iyengar and Greenhouse assume a two-tailed selection process. Therefore, the original file drawer calculations of Rosenthal (1979) are fully consistent with all the formal models in Iyengar and Greenhouse's Section 4, which are used to illustrate their preferred maximum likelihood approach.

The Iyengar and Greenhouse file drawer calculation (based on the assumptions that the null is true but that only results significant in one direction are published) is a worst case calculation. However, it seems to be less realistic than the assumption of a two-tailed selection process because (a) early in the history of a research domain results in either direction are important news and (b) later in the history of the domain,

Robert Rosenthal is Professor, Department of Psychology and Donald B. Rubin is Professor, Department of Statistics at Harvard University, Cambridge, Massachusetts 02138.