DOCUMENT RESUME

ED 450 153 TM 032 343

AUTHOR	George, Carrie A.
TITLE	Problems and Issues in Meta-Analysis.
PUB DATE	2001-02-00
NOTE	18p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (New Orleans, LA, February 1-3, 2001).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Comparative Analysis; *Effect Size; *Error of Measurement; *Meta Analysis; Social Science Research
IDENTIFIERS	Research Replication

ABSTRACT

Single studies, by themselves, rarely explain the effect of treatments or interventions definitively in the social sciences. Researchers created meta-analysis in the 1970s to address this need. Since then, meta-analytic techniques have been used to support certain treatment modalities and to influence policymakers. Although these techniques originally showed great promise, criticisms of the method have developed. This paper explores the problems and issues associated with meta-analysis. These include problems of comparison of different independent variables and different variable constructs and problems that may arise from methodological flaws in some of the studies considered. Clerical errors, misinterpretation of effect sizes, and replicability issues are also concerns that need to be addressed. Missing data and the potential lack of statistical independence among effect size estimates are additional problems. The researcher is urged to take these issues into consideration when deciding whether or not to implement meta-analytic techniques. (Contains 21 references.) (SLD)

Running Head: PROBLEMS AND ISSUES IN META-ANALYSIS

Problems and Issues in Meta-Analysis

Carrie A. George

Texas A & M University

Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, February 1-3, 2001.

Abstract

Single studies, by themselves, rarely definitively explain the effect of treatments or interventions in the social sciences. Researchers created meta-analysis in the 1970s to address this need. Since then, meta-analytic techniques have been used to support certain treatment modalities and to influence policy makers. While these techniques originally showed great promise, criticisms of the methods have developed. This paper explores the problems and issues associated with meta-analysis. The reader is encouraged to take these issues into consideration when deciding whether or not to implement meta-analytic techniques.

Problems and Issues in Meta-Analysis

Since research in the social sciences began, studies have been done to try to understand human behavior. However, when reviewing individual studies, one could easily become more rather than less confused about human behavior. This is because results across individual studies in a given area of inquiry can fluctuate wildly due to sampling error and differences in design quality. For example, one study might support using relaxation training with anxious adolescents while another study might state that this intervention is not useful. With this kind of fluctuation, researchers as early as the 1930s began exploring ways to aggregate results across research studies to better understand what trends were appearing (Kulik & Kulik, 1992). By the 1970s, meta-analytic techniques evolved to deal with this situation (Wolf, 1986).

The current paper serves several purposes. First, the history of meta-analysis will be described. Second, a description of how to do a meta-analysis will be covered. Third, problems and issues surrounding the use of meta-analysis will be explored. Finally, factors that might impede a complete understanding of research are noted. With knowledge of the pros and cons, better meta-analyses can take place.

History of Meta-Analysis

Meta-analysis by definition is "the application of statistical procedures to collections of empirical findings from individual studies for the purpose of integrating, synthesizing, and making sense of them" (Wolf, 1986, p. 5). Meta-analysis may seem commonplace today, but the method has only been widely used since the 1970s. Glass (1976) was the first person to coin the term "meta-analysis". He was not, however, the first person to discuss the need for it. Early in the 20th century, people such as Karl Pearson, W. G. Cochran, and R. A. Fisher mentioned in journals that there was a need to consolidate the literature in a given field. Some of the first uses of meta-analyses were with agricultural studies (Wolf, 1986).

Before Glass discussed meta-analysis in the 1970s, researchers utilized other methods to summarize results across studies (Kulik & Kulik, 1992). Among the other methods are the narrative procedure, the vote-counting technique, the accumulation of probability-values (p-values) across studies, and literature review (Hunter & Schmidt, 1990). The narrative procedure is the oldest technique. In this technique, the "reviewer takes the results reported in each study at face value and attempts to find an overarching theory that reconciles the findings" (Hunter & Schmidt, 1990, p. 468). As can be imagined, this can become quite a cumbersome task when the number of studies reviewed is large. In the voting technique, the researcher simply tallies the statistically significant and non-significant findings across the studies reviewed to determine the general trend (Light & Smith, 1971). This technique can be problematic because of differences in sample size and sampling error variance.

Techniques that use the accumulation of p-values across studies also face the same problems (Hunter & Schmidt, 1990). With regard to using p-values by themselves, Hays (1981) noted, "virtually any study can be made to show significant results at some sample size" (p. 293). Another problem with integrative studies that use only p-values is that they do not consider effect sizes. According to the American Psychological Association style manual (APA, 1994), p-values, are not sufficient to explain the magnitude of the effect in a study because p-values "depend on sample size" (APA, 1994. p. 18). Thompson (1994a) agreed stating that p-values "must (and do) take sample size influences into account" (p. 5). Using effect sizes when comparing studies is more beneficial. Effect size, by definition, is considered to be "the degree of departure from the null hypothesis of a given experiments results" (Standley, 1996, p. 103). It is difficult to compare studies with p-values alone, because of the sample size issue. With effect sizes, better comparisons can be made. The final method mentioned to consolidate findings across studies is the literature review. Literature reviews depend on p-values primarily and the same issues may occur when there are variations in sample size (Cook et al., 1992).

Because there were many problems with the methods mentioned above, new methods were needed. Glass addressed this need. In 1976 Glass suggested that there are three types of research: primary, secondary, and meta-analysis. Primary research is the initial study. Secondary research involves using better methods and statistical techniques to answer the same research question. A meta-analysis, however, is an overarching summary of the primary and secondary studies in the field. Meta-analysis integrates diverse research findings and uses effect sizes as a means to review studies.

Meta-Analytic Procedures

When doing a meta-analysis there are certain procedures that most researchers follow (Lyons, 1998). First, a problem must be identified, such as "Does cognitive-behavioral therapy work?" Then, a literature search begins. How the literature search is done is critical to the results gained from the meta-analysis.

One problem that could occur when doing a literature search relates to the "file drawer problem" (Rosenthal, 1979). In the file drawer problem, many studies are lost to the researcher doing a meta-analysis. The studies that fall into the file drawer are ones that do not show statistically significant p-values and subsequently are not accepted by peer reviewed journals. Besides research that is not statistically significant, dissertations and master's thesis also are important places for the meta-analyst to look (Wolf, 1986). It is only with an exhaustive search of the file drawers and graduate papers can the researcher feel that they have a good grasp of what is going on in the field.

When literature that fits the problem question are identified, the studies are read and coded according to different methodological criteria such as "research design factors, sample characteristics, type of dependent variable" (Lyons, 1998, p. 13).

The next step is to determine the effect sizes of each study. There are many different measures of effect size (Kirk, 1996; Snyder & Lawson, 1993). According to Standley (1996), "determination of an appropriate measure of effect size is dependent on the data, tests, and statistical models used in the studies included in the meta-analysis" (p. 103). Effect sizes can be in the form of descriptive or inferential statistics (Standley, 1996). The effect size created by Glass, McGaw, and Smith (1981) is a descriptive statistic and it can be calculated as a z-score. The formula for this effect size, also called the standardized mean difference, is:

$$ES = \frac{X_E - X_C}{SD_C}$$

To calculate the effect size in this formula the control group mean (X_C) is subtracted from the experimental group mean (X_E). This number is then divided by the standard deviation (SD_C) of the control group.

Cohen (1988) offers a similar formula for the effect size that he refers to as d. Cohen's d is an inferential statistic and it is similar to that of Glass and his colleagues. The formula is:

$$d = \frac{M_T - M_C}{S_P}$$

In this case the mean of the control group (M_C) is subtracted by the mean of the experimental group (M_T). The value is then divided by the pooled standard deviation (S_P) estimate. Some researchers then use one of Cohen's formulas is to convert d to r (a correlation coefficient) or r^2. The r^2 value is the "proportion of total variance in the combined populations under study" (Standley, 1996, p. 103).

After the effect sizes for each study are determined, an overall mean effect size is calculated (Lyons, 1998). This "typical" effect size must then be explained. An example of this would be to state that the "average effect size of .95 indicates a 75 percent probability that a randomly selected response in the treatment group exceeds a randomly selected response in the control group" (Cook et al., 1992, p. 319). Finally, the last stage of a meta-analysis is to determine the overall meaning behind the results.

Problems in Meta-Analysis

When meta-analysis came into being, it seemed quite exciting. In fact, there was a great increase in the use of meta-analysis since the 1970s. According to Lyons (1998), there were only 51 studies that used the term meta-analysis before 1983, but between 1983 and 1990 there were over 858 journal articles that used the term. It is certainly a concept that has taken off. As meta-analysis grew in popularity, so too did criticisms grow.

Apples and Oranges

An overall concern with meta-analyses is that meta-analysts are comparing apples to oranges (Cook et al., 1992; Hunter & Schmidt, 1990; Wolf, 1986). This criticism has usually been applied to "Glassian" meta-analysis. The criticism involves the basic idea that meta-analysis is combining and comparing "different independent variables and different dependent variables...[and that] the independent and dependent variable constructs vary across studies in the same meta-analysis" (Hunter & Schmidt, 1990, p. 516). Hunter and Schmidt (1990) stated that it might be fine to do this because "any set of numbers can be compared, averaged, or otherwise analyzed without logical contradiction" (p. 516). Basically, their argument is that the purpose of meta-analysis is to analyze results and not to analyze studies. They feel it is purely mathematical. While this may be true, it still does not help the validity of meta-analyses if all the studies reviewed are completely different. Meta-analysis should involve more thinking and less math.

Other Methodological Errors

One of the major concerns is that many meta-analyses take all studies into consideration, even despite methodological errors (Wolf, 1986). Well-designed studies are analyzed along with poorly designed studies. Validity issues can also be a concern with the poor research design. For example, one study might use measures that yield reliable and valid dependent variable scores, while another might select variable measures with scores that are unreliable or invalid. In addition, the subject pool in different studies is questionable. Study A might be a study of play therapy techniques with five year old depressed females in a clinical setting, while study B might use play therapy with 8 year old hyperactive males in a school setting. It would be difficult to come to the conclusion that play therapy is effective for children, when there are other factors that might be involved. Of course, all these study variations can be statistically controlled in the meta-analysis by coding and analyzing the influences of these study features on effect sizes.

Clerical Errors

Other problems that could occur with meta-analysis relate to clerical errors (Wolf, 1986). Today at least 12 journals now require the reporting of effect sizes to have articles published. For example, the editorial policies of the journal Educational and Psychological Measurement state that "Authors reporting statistical significance [are] required to both report and interpret effect sizes" (Thompson, 1994b, p. 845). When the material pulled for the meta-analysis does not have the effect size reported, the researcher doing the meta-analysis must calculate those values by themselves. Many times careless errors can be made in this stage of the meta-analysis. Some researchers have others coding the data from the studies. The many people involved may lead to a higher chance of error (Wachter & Straf, 1990). It is imperative that all journals require the reporting of effect sizes, because of the importance of the data derived from them and for the purpose of future meta-analyses that may be done (Wilkinson & APA Task Force on Statistical Inference, 1999).

Misinterpreting Effect Sizes

According to Cook et al. (1992), the effect sizes obtained can also be misinterpreted. Effect sizes should be understood in the context of the field. As Cohen (1969) stated, effect sizes "may not be reasonably descriptive in any specific area. . . .Thus what a sociologist may consider a small effect may well be appraised as medium by a clinical psychologist" (p. 278)

Replicability Issues

Regarding replicability, there is a concern that one meta-analysis done in an area may not be the same as another one done in the same area. Problems begin with the initial literature search. Due to differing views and methods, the same studies may not be reviewed (Wolf, 1986). For example, one researcher may decide to search for literature on only one database (e.g., ERIC). Another researcher may search on that database and in unpublished articles and dissertations. The effect sizes from the less comprehensive search are likely to be an underestimate of what is really happening in the population (Cook et al., 1992). In addition, if the first researcher were to publish the results, then there would be a greater chance for a "Type I publication bias error [due to] more positive results than is really the case" (Wolf, 1986, p. 38). If both meta-analyses were published they might not even agree. This would be difficult to explain to legislators or other consumers. If policy makers were to view this discrepancy, differences would be difficult to explain.

Missing Data

When doing meta-analyses, primary studies may vary so much that the researcher is left with gaps while coding the studies. Studies also might not report enough information to calculate the effect size (Wachter & Straf, 1990). Missing data in primary studies is usually handled in different ways. Techniques used in primary studies include listwise deletion, pairwise deletion, mean substitution and regression estimation (Cool, 2000). In meta-analysis, some researchers tend to ignore the problem or impute (replace) all missing values with zero. Doing the latter may underestimate the overall effect size and it can also create less variance among the effect sizes.

According to Wachter and Straf (1990), the best technique for handling missing data in meta-analysis is "extract[ing] from the study any available information about the effect size" (p. 17). The authors stated that at least the direction of the effect could be figured out from the data given. The meta-analytic researcher must have a method to deal with missing data because many studies are likely to have missing information. Overall effect sizes cannot be calculated without all the information, so some method to fill in the gaps should be chosen. The researcher must be wise in choosing a method.

Statistical Dependence

A final problem to be discussed is "the potential lack of statistical independence among effect size estimates" (Wachter & Straf, 1990, p. 24). Dependence among the effect sizes can be very difficult to accomodate. Dependence can occur when effect sizes for the primary studies are being calculated. The researcher might feel it necessary to code and calculate all effect sizes possible in a study to avoid losing information. When multiple effect sizes are calculated from within a single study and then added to the general pool of studies, a single study can skew the overall effect size. This also tends to increase the Type I error rate (Wolf, 1986). Another dependence problem occurs when the same primary researcher (or research team) does many of the studies chosen in the meta-analysis. This can be troublesome because there may be less variation in this group of studies than those done by independent researchers. This type of dependence may also cause problems in the estimation of the overall effect size (Wachter & Straf, 1990).

Conclusion

When looking at the problems associated with meta-analysis, one may begin to think that this method is no better than previous methods. This is not the argument here. The argument is that the researcher should think through these issues when conducting a meta-analysis. It is also important that policy makers know about some of the methodological problems associated with meta-analyses. Legislation should not be based on faulty research. Improving meta-analysis is the answer, not blindly continuing to do the same old techniques. It is possible to improve on the current techniques, according to Wolf (1986):

> better meta-analyses have. . .gone on to explore some of the method factors, some of the populations and settings, and some of the treatment variants that influence the size of the effect. But exploration of such contingency variables is rarely systematic. (p. 14)

A researcher's most important skill is the ability to think. By thinking through the problems that might occur and trying to fix them, better meta-analyses can be done. As Green and Hall (1984) stated, "Data analysis is an aid to thought, not a substitute" (p. 52).

References

American Psychological Association. (1994). Publication manual of the American Psychological Association (4th ed.). Washington, DC: Author.

Cohen, J. (1969). Statistical power analysis for the behavioral sciences. New York: Academic Press.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cook, T. D., Cooper, H., Cordray, D. S., Hartmann, H., Hedges, L. V., Light, R. J., Louis, T. A., & Mosteller, F. (1992). Meta-analysis for explanation. New York: Russell Sage Foundation.

Cool, A. L. (2000, January). A review of methods for dealing with missing data. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, TX.

Glass, G. V (1976). Primary, secondary and meta-analysis of research. Educational Researcher, 5, 3-8.

Glass, G. V, McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage.

Green, B. F., & Hall, J. A. (1994). Quantitative methods for literature reviews. Annual Review of Psychology, 35, 37-53.

Hays, W. L. (1981). Statistics (3rd ed.). New York: Holt, Rinehart and Winston.

Kirk, R. (1996). Practical significance: A concept whose time has come. Educational and Psychological Measurement, 56, 746-759.

Kulik, J., & Kulik, C. L. (1992). Meta-analysis: Historical origins and contemporary practice. In B. Thompson (Ed.), Advances in social science methodology (Vol. 2, pp. 53-80). Greenwich, CT: JAI Press.

Light, R. J., & Smith, P. V. (1971). Accumulating evidence: Procedures for resolving contradictions among different research studies. Harvard Educational Review, 41, 429-471.

Lyons, L. C. (1998). Meta-analysis: Methods of accumulating results across research domains [On-line]. Available Internet: www.mnsinc.com/solomon/MetaAnalysis.html

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. Psychological Bulletin, 86, 638-641.

Snyder, P. & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. Journal of Experimental Education, 61, 334-349.

Standley, J. M. (1996). A brief introduction to meta-analysis. Journal of Research in Music Education, 44(2), 101-104.

Thompson, B. (1994a). The concept of statistical significance testing. Measurement Update, 4(1), 5-6. (ERIC Document Reproduction Service No. ED 366 654)

Thompson, B. (1994b). Guidelines for authors. Educational and Psychological Measurement, 54, 837-847.

Wachter, K. W., & Straf, M. L. (1990). The future of meta-analysis. New York: Russell Sage Foundation.

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. American Psychologist, 54, 594-604.

Wolf, F. M. (1986). Meta-analysis: Quantitative methods for research synthesis. Beverly Hills, CA: Sage.

TM032343

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE
(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: PROBLEMS AND ISSUES IN META-ANALYSIS	
Author(s): CARRIE A. GEORGE	
Corporate Source:	Publication Date: 2/1/01

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] ← Sample sticker to be affixed to document

Check here
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
CARRIE A. GEORGE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document → [ ]

or here
Permitting reproduction in other than paper copy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: Carrie George	Position: RES ASSOCIATE
Printed Name: CARRIE A. GEORGE	Organization: TEXAS A&M UNIVERSITY
Address: TAMU DEPT EDUC PSYC COLLEGE STATION, TX 77843-4225	Telephone Number: (409)
	Date: 1/20/01

OVER