DOCUMENT RESUME

ED 237 895 CG 017 150

AUTHOR Champney, Timothy F.; Schulz, E. Matthew
TITLE A Reassessment of the Effects of Psychotherapy.
PUB DATE 6 May 83
NOTE 39p.; Paper presented at the Annual Meeting of the Midwestern Psychological Association (55th, Chicago, IL, May 5-7, 1983).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Counseling Effectiveness; Counselor Role; *Effect Size; Etiology; Meta Analysis; *Nondirective Counseling; *Research Methodology; Statistical Bias; *Transactional Analysis

ABSTRACT

Smith et al. (1980) analyzed 475 psychotherapy studies and concluded that individuals receiving treatment were better off than 80 percent of the untreated control groups. These studies were criticized on methodological grounds, particularly for failing to enable calculation of an index of effect size. To address these methodological issues, 20 published studies cited in the Smith study, from two treatment domains, i.e., the effectiveness of client-centered therapy (N=17) and transactional analysis (N=3), were coded according to type of therapy, allegiance of investigator, diagnosis, estimation methods, comparison treatment, subject solicitation, treatment center, and source of means. Effect size and population effect size were computed according to the methodology in the original study. An analysis of the results showed a lack of homogeneity of effect size estimates. The success of client-centered therapy and transactional analysis appeared to depend on investigator bias, location of treatment, diagnosis, and presence of a comparison treatment. Specifically, these two treatment methods appeared to be most successful when applied to problems that occur in academic settings. Caution is recommended in generalizing the results of Smith's study, which overly represented academic problems and educational counselors compared to those in medical settings. (BL)

A Reassessment of the Effects of Psychotherapy

Timothy F. Champney
Department of Behavioral Sciences
The University of Chicago

and

E. Matthew Schulz
MESA Group
Graduate School of Education
The University of Chicago

Rehabilitative Research & Development Center
U.S.V.A. Hospital, Hines, IL

Paper presented at the Midwestern Psychological Association Annual Meeting, Chicago, Illinois, May 6, 1983.

INTRODUCTION

The question of the effectiveness of psychotherapy has been examined in many studies. Several authors have attempted to synthesize the results of these studies quantitatively (Smith and Glass, 1977; Smith et al., 1980) through the use of meta-analysis (Glass, 1976). Based on the results of 375 studies, Smith and Glass (1977, p.754) claimed that "the average client receiving psychotherapy was better off than 75% of the untreated controls." Smith et al. (1980) claimed that individuals receiving psychotherapy in any form were better off than 80% of the individuals in untreated control groups after analyzing an additional 100 studies (475 studies total).

These conclusions soon drew sharp criticism. Several authors (for example Eysenck, 1978) ojected to the quality of the studies being synthesized, particularly the nature of the control groups. The meta-analysis methods used by Smith et al. (1980) have also been criticized. Landman and Dawes (1982) have pointed out that Smith and Glass (1977) treated 833 measures of effect size from the 375 studies as independent data points for purposes of computing overall average effect sizes and in order to regress effect sizes on study characteristics. To correct this problem, Landman and Dawes (1982) computed an average effect size for each study from a sample of the original studies. They found that the original conclusions were not substantially changed by this procedure. In their discussion, Landman and Dawes (1982) pointed out, however, that 26% of the studies that they sampled "failed to report one or more sets of results in a form ena-

biing calculation of an index of effect size." The authors implied that the letters "NS" were often substituted in statistical tables in place of sufficient information for the calculation of effect sizes. As they mention, Greenwald (1975) has noted two related sources of bias -- the failure of investigators to report nonsignificant results and the failure of journals to publish them. Using simulation methods, Lane and Dunlap (1978) have shown that editorial decisions to suppress publication of nonsignificant results may lead to large exaggerations of effect sizes in meta-analyses.

Hedges (1982) has raised several additional methodological issues that lead us to question the conclusions of Smith et al. (1980). The possibility that a variety of non-homogeneous measures, that are not linearly equatable (like apples and oranges) have been lumped together concerns us particularly. Hedges (1982) proposes a test of homogeneity of effect sizes. Hedges and Olkin (1983, in press) have developed methods for partitioning effect sizes into homogeeous groups. And, Hedges (1983, in press) has developed a method for testing the homogeneity of effect sizes between and within groups (as in the analysis of variance).

One further methodological objection to the Smith et al. (1980) conclusions must be raised. Cronbach (1982), among others, has pointed out the need to examine the representativeness of a sample with regard to the population that one hopes to generalize to in evaluating the effects of a treatment. As pointed out by Hedges (1982), the Glass (1976) meta-analysis procedure ignores issues of inference about

population treatment effects completely. While Smith et al. (1980) make wide ranging inferences about the general (population) effects of psychotherapy from the results (sample), they have failed to utilize sound statistical inference and probability sampling procedures that make valid statistical inference possible. Practically speaking, we must ask whether or not these 475 studies are representative of all of psychotherapy.

To address the above issues we selected 20 published studies cited in Smith et al. (1980) from two treatment domains, client-centered therapy and transactional analysis. The overlap of these studies with those catalogued by the National Library of Medicine was examined. We computed an average effect size for each of the 20 studies and estimated the population effect size by both the methods of Glass (1976) and Hedges (1982). The homogeneity of effect sizes was tested and rejected. Following the overall test we partitioned the studies into homgeneous groups. Study characteristics within the subgroups were examined. The issue of bias due to editorial decisions to publish only significant results was examined in Appendix B. Maximum likelihood techniques for truncated data (Cohen, 1950) were applied to estimate the population effect of psychotherapy in a worst possible case, assuming all studies with significant results were selected due to their statistical significance. We estimated, at worst, that editorial decisions could have inflated the results by as much as 60%.

METHODS

Studies concerning the effects of client-centered therapy or transactional analysis were selected from the Smith et al (1980) bibliography for coding. The selection was based on the presence of keywords (such as "client-centered," "nondirective," or "transactional") in the titles indicating one of these orientations. We also selected studies if the investigator is known to be of one of the above orientations. The National Library of Medicine data base (MEDLINE) was searched for the years 1966 through 1977 to establish the thoroughness of our coverage of the literature. Due to cost and time limitations we did not include dissertations or unpublished papers in our analyses.

The selected studies were located in university libraries. We then coded study characteristics on an abreviated version of the Smith et al (1980) coding form. Those characteristics reported as predictive of psychotherapy effect size were retained on the form. Any information needed for effect size computation was also coded. The abreviated form appears as Appendix A.

A further modification of coding procedures was necessary due to the relatively small number of studies in our sample. The categories in the coding form of Appendix A were often not used or accounted for too few studies. We therefore attempted to combine similar categories and to reduce the number of study characteristics coded. 'Control Treatment' was reduced to two categories - 'no treatment' or 'any

other form of treatment'. The 'solicitation of clients was reduced from 5 categories to two - 1) presentation in response to advertisement, or autonomous presentation, or solicited by experimentor and 2) committed or referred. We did not classify the outcome measures for the studies, because we wished to generalize our analysis to all studies employing either client-centered therapy or transactional analysis, and to all outcome measures typically employed with these types of studies. In all, eight characteristics were coded, with two to five categories each. Table 1 shows this modified coding scheme.

In most cases the effect sizes are computed according to the methods given in Glass et al. (1981). This usually involves the conversion of independent or paired t statistics to effect size estimates. When more than one effect was reported in a study we computed the average t statistic in order to avoid the problem of non-independence. In some instances proportions (i.e. proportion "improved" following therapy) or nonparametric statistics were converted to effect size estimates by referring to a normal probability table.

The effect sizes were assumed to have the statistical properties of a standardized mean difference, g_i, where,

$$g_i = \frac{\bar{Y}_i^E - \bar{Y}_i^C}{S_i}$$

In this equation, $\bar{Y}_i^E$ and $\bar{Y}_i^C$ are the experimental and control group means on the outcome variable, or variable-combination, of a study and S_i is the pooled standard deviation, i.e.,

$$S_p^2 = \frac{S_C^2 N_C + S_E^2 N_E}{N_C + N_E}$$

Hedges (1980) has noted, however, that the true effect size, δ, estimated by g_i is a noncentral t-variate. The unbiased estimate of δ is g_i', where

$$g_i' = g_i \times c(m_i)$$

The value, 'm_i' is the sum of experimental and control group sizes, minus two, for study 'i'. The correction, '$c(m_i)$' is approximated to a maximum error of 0.007 at $m_i = 2$ (Hedges, 1981) by the equation:

$$c(m_i) = 1 - \frac{3}{4m_i - 1}$$

Unbiased effect sizes, g_i', were thus calculated for all studies from the initial effect size estimates.

Hedges (1982) procedures were then used to test the hypothesis that all of the studies shared a common effect size. This procedure is based on a general one-way ANOVA comparison of effect sizes within and between groups of studies. The FORTRAN program shown in Appendix C was used to perform the effect size ANOVA according to the computational equations given in Hedges (1982). The variance, σ, of an effect size, δ, is estimated for each effect size with the equation,

$$\hat{\sigma}^2(\delta_{ij}) = \frac{N_{ij}^E + N_{ij}^C}{N_{ij}^E N_{ij}^C \rho} + \frac{g_{ij}'^2}{2(N_{ij}^E + N_{ij}^C)}$$

where 'ij', corresponds to the i'th study in the j'th category. The reliability, ρ, of the outcome measure is set to 1 for all effect

sizes. Other critical values calculated in the analysis of variance are $g'_{..}$, the weighted estimate of the common effect size for all studies, and $g'_{.j}$, the weighted estimate of the common effect size for all studies within a given category, 'j'. The final values calculated by the ANOVA program in Appendix C are H_T, the sum of squared deviations of effect sizes from the weighted common effect size; H_W, the within groups sum of squared deviations of effect sizes from their weighted group effect size; and H_B, the between groups sum of squares. H_T, H_W and H_B are distributed as Chi Square statistics with m, p-1, and m-p degrees of freedom respectively. 'M' equals the number of effect sizes, 18, and 'p' equals the number of categories into which studies are grouped.

A cluster analysis was also performed to divide the studies into homogenous groups, with a common effect size for each group. This analysis converts the effect sizes into standard normal deviates and compares the difference between two given normal deviates, U_i and U_j to a critical difference corresponding to a given probability level for observing a difference as large or larger. The FORTRAN program used for this analysis is shown in Appendix D. The program implements the procedures outlined by Hedges and Olkin (1983, in press). A variance-stabilizing transformation, h_i, is first calculated for each effect size according to the equation,

$$h_i = \sinh^{-1}(g'_i/2\sqrt{2})$$

When experimental and control groups have unequal sizes, $2/\sqrt{2}$ is replaced by

$$\sqrt{(2 \times (1 + N_E/N_C)^2)/(N_E/N_C)}$$

The variance stabilized transformations are then transformed to standard normal variates with the equation,

$$U_i = 2\ n[h_i - \bar{h}]$$

where 'n' is the square mean root of the sample sizes of k-t studies; k = 18, and t = the number of studies with 'outlying' sample sizes, i.e., greater than 100 in this case. The U_i are assumed to be normally distributed around the mean, $\bar{h}$, where

$$\bar{h} = \frac{\sum_{i=1}^{k} n_i h_i}{\sum n_i}$$

The value, n_i is the average sample size for study 'i' - $(N_i^E + N_i^C)/2$.

Both disjoint and overlapping cluster analysis was performed. In disjoint clustering, the difference between adjacent U_i's is compared to significant gap sizes between standard normal order statistics for a sample of size 18 and a significance level ρ. Disjoint clustering compares the difference between the highest and lowest U_i of a specified cluster to critical values for Bonferroni multiple comparisons of the range of standard normal variates for significance level, ρ. The critical values were obtained from Tables 1 and 2 of Hedges and Olkin (1983, in press), and linear interpolation was performed when the exact sample sizes were not given. Significance levels were initially set at 0.05, and increased to 0.2.

RESULTS

The Smith, et al. (1980) bibliography yielded sixteen published studies employing client-centered therapy and three published studies employing transactional analysis. One additional study that deviated somewhat from the client-centered protocol but which used client-centered outcome measures was also retained for coding. The MEDLINE search yielded 299 citations of articles on the above types of therapy. Some of these appeared to be papers on therapeutic technique or reviews rather than experimental investigations of treatment effects. Of the 299 articles located by MEDLINE, only 2 were also cited by Smith et al. (1980). Nine had authors also cited by Smith, et al., but different titles. Ten of the articles obtained from the Smith et al (1980) bibliography have authors who were not located through MEDLINE.

The number of studies in each category of the coded study characteristics is shown in Table 1. These numbers are enclosed in parentheses. Thus, fifteen of the studies used no treatment as the 'type of comparison treatment' and three studies gave the comparison group some other type of therapy (usually behavioural). The allegiance of the experimentor was considered 'favorable' to the experimental treatment in fourteen of the studies, and 'unfavorable' or 'indifferent' in four. The other study characteristics in Table 1 can be similarily interpreted. In general, a large proportion of studies centered on academic or character problems and were located in academic settings. The estimation methods usually were based on

Table 1: Coded characteristics of studies, categories, category codes, frequencies.

TYPE OF THERAPY

1 = Transactional analysis (3)
2 = Client-centered (15)

ALLEGIANCE OF INVESTIGATOR

1 = Favorable (14)
2 = Indifferent or Unfavorable (4)

DIAGNOSIS

1 = Academic difficulties (4)
2 = Character disorder (5)
3 = Neurotic (3)
4 = Psychotic (5)
5 = Normal (1)

ESTIMATION METHODS

1 = T-values (9)
2 = Mean-difference/S.D. (4)
3 = Chi Square (3)
4 = Nonparametrics (1)
5 = Other (1)

TYPE OF COMPARISON TREATMENT

1 = No treatment (15)
2 = Other treatment (3)

SOLICITATION OF SUBJECTS

1 = Autonomous or Advertisement or Solicited by experimentor (12)
2 = committed or referred (5)

TYPE OF TREATMENT CENTER

1 = School (7)
2 = Detention institution (2)
3 = Private outpatient (2)
4 = Inpatient (5)
5 = College mental health (2)

SOURCE OF MEANS

1 = Unadjusted post-test (8)
2 = Covariance-adjusted (1)
3 = Residual gains (2)
4 = Pre-post differences (4)
5 = Other (3)

Table 2: Studies and their characteristics, listed in order of effect size.

No.	Authors and Date	Exp.	Comp.	Alleg.	Solicit	Diag.	Locat.	E.Meth.	S.Means
1	Lazarus, 1966	1	1	1	1	3	3	3	5
2	Ashby, et al, 1957	1	1	2	1	3	5	5	1
3	Alper & Kranzler, 1970	1	2	1	2	1	2	2	1
4	Jesness, 1975	2	1	1	2	2	2	1	2
5	Andrews, 1971	1	2	1	1	5	1	2	4
6	Rogers, et al, 1967	1	2	2	1	4	4	2	1
7	Roth, et al, 1967	1	2	2	2	1	1	1	1
8	Carkhuff & Truax, 1965	1	2	2	1	4	4	3	5
9	Truax & Wittmer, 1971	1	2	2	-	4	4	1	1
10	Seeman, et al, 1964	1	2	2	1	2	1	1	4
11	Varble & Landfield, 1969	1	2	2	1	3	5	1	1
12	Truax, et al, 1966	1	2	2	2	2	4	1	3
13	Truax, 1970	1	2	2	1	4	4	1	5
14	Truax & Tourney, 1971	1	2	2	1	2	1	3	5
15	Coche & Douglas, 1977	2	2	2	1	4	3	4	3
16	Dorfman, 1958	1	2	2	2	2	1	1	4
17	Sheldon & Landsman, 1950	1	2	2	1	1	1	1	1
18	Beckstraud, 1973	2	2	2	1	1	1	2	4

t-statistics or were directly calculated from the mean differences and pooled standard deviations according to the equation given earlier for effect sizes. Means were most often based on unadjusted post-test measures.

Other charactaristics of the studies were not coded, but are noteworthy. Of the 20 articles coded, 95% are from journals and 5% appear as books. Cases handled by counseling psychologists as opposed to clinical psychologists and psychiatrists appear to be over-represented. Twenty percent involve juvenile delinquency problems and 20% academic problems. Forty percent of the studies use subjects under age 18 and an additional 25% use college age persons. Seventy percent used objective tests.

Table 2 shows the coded characteristics of the studies, individually. Studies with negative effect sizes are listed first, while studies with high, positive effect sizes are listed last. With this order in mind, it is possible to see that unfavorable experimentor bias, a diagnosis of a neurosis, the use of comparison treatments and location in a detention institution tend to be associated with each other and to lead to negative, or lower effect sizes. At the other end of the continuum, the two studies with the highest effect sizes are unique in that they are both concerned with academic difficulties and are located in a regular school setting. For these two studies, as well as all other studies with positive effect sizes, the investigator's bias was favorable to the experimental treatment.

Table 3: Sample sizes, biased and unbiased effect sizes, the effect size variance, the variance-stabilizing transformation, and the standard normal variate transformation.

No.	Authors and Date	N^E	N^C	g_i	g_i'	$\sigma^2(g_i)$	TRANS	S.N.V.
1	Lazarus, 1966	25	25	-.470	-.463	.082	-.164	-3.33
2	Ashby, et al, 1957	23	23	-.185	-.182	.087	-.064	-2.35
3	Alper & Kranzler, 1970	9	9	-.018	-.017	.222	-.006	-1.78
4	Jesness, 1975	427	482	.045	.045	.004	0.016	-1.56
5	Andrews, 1971	23	23	.110	.108	.087	0.038	-1.34
6	Rogers, et al, 1967	24	24	.156	.154	.084	0.054	-1.18
7	Rothe, et al, 1967	52	52	.365	.362	.039	0.128	-0.45
8	Carkhuff & Truax, 1965	70	74	.440	.438	.028	0.155	-0.19
9	Truax & Wittmer, 1971	16	16	.464	.453	.128	0.160	-0.14
10	Seeman, et al, 1964	8	8	.515	.489	.258	0.173	-0.01
11	Varble & Landfield, 1969	36	35	.530	.524	.058	0.185	0.11
12	Truax, et al, 1966	30	30	.630	.622	.070	0.220	0.45
13	Truax, 1970	16	16	.652	.636	.132	0.225	0.50
14	Truax & Tourney, 1971	30	20	.771	.759	.089	0.263	0.88
15	Coche & Douglas, 1977	25	29	.806	.795	.080	0.280	1.05
16	Dorfman, 1958	17	17	.837	.818	.128	0.289	1.13
17	Sheldon & Landsman, 1950	9	10	1.265	1.212	.253	0.428	2.50
18	Beckstraud, 1973	33	38	1.592	1.575	.074	0.555	3.76

Table 3 shows the principle values in the analysis of variance and clustering procedures. Sample sizes ranged from 8 to 482. The effect sizes, corrected for bias, ranged from -.463 to 1.575. The range was slightly greater without correcting for bias. The weighted average of biased effect sizes was .252. The weighted average of unbiased effect sizes was .249. The unweighted average of biased effect sizes, corresponding to Glass's (1976) method, was .472. The weighted average takes the variance of the estimate of each effect size into account, which in turn is a function of sample size. The largest variance was .258 for groups of size 8 and the smallest was .004 for groups of size 450. The variance stabilized transformations of effect sizes ranged from -.164 to .555, with a weighted mean of .16. The standard normal variates are centered around a mean of 0.

Table 4: Analysis of variance for homogeneous effect sizes

CLASSIFICATION	NUMBER OF CATEGORIES	H_T	H_W	H_B
NONE	1	61.5[6]		
TYPE OF THERAPY	2		57.2[6]	4.5
TYPE OF COMPARISON TREATMENT	2		28.3[5]	33.3[3]
ALLEGIANCE OF INVESTIGATOR	2		30.4[5]	31.2[3]
SOLICITATION OF SUBJECTS	2		49.7[6]	11.9[2]
DIAGNOSIS	5		40.7[6]	20.9[3]
TYPE OF TREATMENT CENTER	5		33.6[6]	28.8[2]
ESTIMATION METHODS	5		49.6[6]	12.0[2]
SOURCE OF MEANS	5		32.8[6]	28.8[3]
DISJOINT CLUSTERS	3		33.0[3]	28.6[2]
OVERLAPPING CLUSTERS				
CLUSTER 1	1	23.3[4]		
CLUSTER 2	1	32.7[5]		
CLUSTER 3	1	31.4[5]		
COMBINED CLUSTERS	3	61.5	10.8[1]	50.9[3]
STUDIES 1-5,6-16,17-18	3		8.9[1]	52.6[3]

[1] Homogeneous groups P > 0.8
[2-3] Between group differences [2] P < .01, [3] P < .001
[4-6] Nonhomogeneous groups [4] P < .05, [5] P < .01, [6] P < .001 or less

Table 5: Grouping of effect sizes according to cluster analysis.

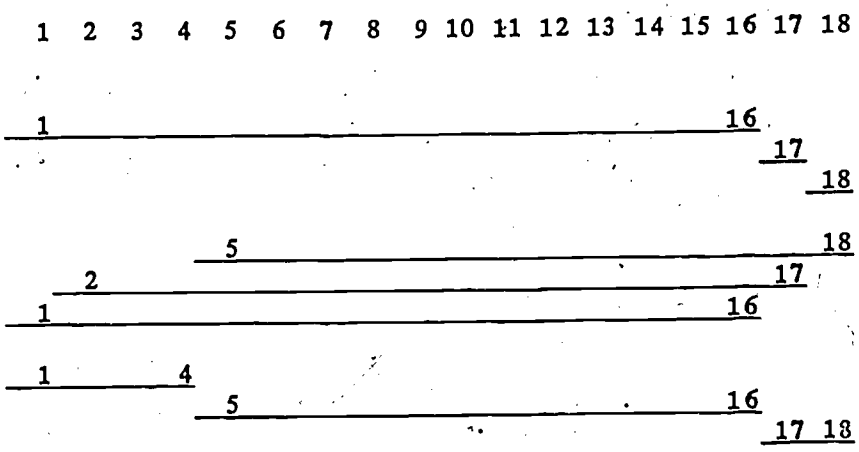

Table 4 shows the analysis of variance results for various groupings of studies, based on effect sizes corrected for bias. The analysis of variance showed that the effect sizes were not homogeneous (H_T = 61.5, $p < .00001$) None of the classifications used in Table 1 lead to groupings with homogeneous effect sizes. The lowest within groups sum of squares was found for studies grouped according to whether the comparison group recieved an alternative type of treatment. Groupings according to the allegiance of the investigator also produced a relatively low sum of squares within groups. Groupings according to the solicitation of subjects and estimation methods were noteably heterogeneous. Significant between-group differences were found for all classifications except 'type of therapy'. We did not estimate aver effect sizes for the categories of Table 1, however, because of the lack of effect size homogeneity.

Table 4 also shows the results of the analysis of variance performed on the study groupings obtained from the cluster analysis. The clusters are illustrated in Table 5. The disjoint clusters, separated by a gap length significant to the 0.2 level, consist of studies 1-16, 17, and 18. The study numbers are taken from Tables 2 and 3, and reflect increasing positive effect sizes. The overlapping clusters, formed by critical range sizes at a ρ of .02, consist of study numbers 1-16, 2-17 and 5-18. None of the groupings formed by the disjoint clustering or the overlapping clustering alone were homogeneous, according to the analysis of variance. The hypothesis of homogeneity was rejected at the .05 to .01 level of confidence. However, we used the results of both cluster analyses to form three groups of studies -

1-4, 5-16 and 17-18 which proved to have a small within groups sum of squares of 9.2, $p > .8$, and a large between groups sum of squares of 54.1, $p < .0001$. The hypothesis of homogeneity was not rejected. For better interpretability, in view of the association of characteristics with effect size magnitude brought out by Table 2, the fifth study was added to the lower group giving studies 1-5, 6-16 and 17-18 in the three groups. Table 4 shows that this new arrangement further decreased the sum of squares within groups and increased the sum of squares between groups.

The average effect sizes for the three groups were 0.013 for studies 1-5; 0.52 for studies 6-16; and 1.52 for studies 1.52.

The major characteristics of studies 1-5 were investigator bias unfavorable to the experimental treatment and/or the use of alternative treatments for the comparison group. Studies in the intemediate group seemed to have no unique characteristic. However, the investigator bias was uniformly favorable to the experimental treatment in the intermediate studies. The studies with the highest effect sizes, as already mentioned were concerned with minor academic problems in a normal school setting.

DISCUSSION

One of the most striking features in our results is the lack of homogeneity of effect size estimates. The hypothesis of homogeneity of effect sizes for our 18 studies is rejected at $p < .00001$. Since

our effect sizes represent a subset of the even more wide ranging sample of studies used in the meta-analysis of Smith et al. (1980), we expect that if their data were reanalyzed using our techniques the lack of homogeneity would be even more apparent.

Given that the studies are non-homogeneous we asked if the studies could be partitioned into interpretable subgroups? Through the application of the clustering methods of Hedges and Olkin (1983, in press) we obtained 3 homogeneous groups. For better interpretability we moved the study with the smallest effect size in the intermediate cluster to the cluster of studies with the smallest effect sizes. The resulting three groups are even more homogeneous and can be characterized according to investigator bias, location of treatment, diagnosis, and presence or absence of a comparison treatment. The success of client-centered therapy and transactional analysis therefore appears to depend substantially on these, and possibly other factors.

Specifically, client centered therapy and transactional analsysis appear to be most successful in the treatment of problems that occur in academic settings. Of our three homogeneous groups, the one with the highest effect size, 1.52, consisted of studies that used volunteer subjects with minor academic problems (a diagnostic category omitted in Smith et al., 1980). Psychotherapy in general, however, may be more effective in academic settings. If we divide the Smith et al. (1980, p 118) effect sizes into academic (68%) and nonacademic (36%) settings, we find an average effect size of .91 in academic and .74 in nonacademic settings. From the homogeneous grouping at the

other extreme of effect size, we find that client centered therapy and transactional analysis offers little or no advantage over other forms of therapy and tend to be reported to have no effect by investigators nonallegiant to these types of therapy. The 5 studies with the lowest average effect size of 0.09, were reported by investigators allegent to other forms of therapy or of unknown allegiance, and usually involved alternative treatments for the comparison groups. We note that Smith et al. (1980, p.121) report smaller effect sizes for studies conducted by experimenters with opposed or unknown allegiance.

Our findings differ in several respects from those of Smith, et al (1980). Smith et al (1980) reported an average effect size of .62 from 150 client-centered effect sizes and an average of .67 from 28 transactional analysis effect sizes. Weighting these averages by the number of effect sizes in each class we obtain an average effect size of .63 from the Smith et al. (1980) results. We assume this average reflects all comparisons, including comparisons with other types of treatment. Our estimate by the same methods, based on all 18 effect sizes in our review is only .47. Leaving out comparisons with other forms of treatment, we obtain an average effect size of only .55. Our finding a lower average effect size, using the same methods for calculation, may be due in part to our 18 studies giving the academic setting less representation than did Smith, et al, as a whole. Our use of different methods to obtain an average effect size brings up another issue. Hedges' (1980) methods take the variance of the effect size estimates, and sample sizes, into account. Thus the large study, with over 900 subjects, and a near zero effect size receives much more

weight than it would by Glass's method of simply averaging effect sizes. Our weighted average of all eighteen effect sizes is .249, using unbiased effect size estimates. We cannot, however, take .249 to be a generalizable estiamte given the results of our analysis of variance for effect size homogeneity. Nor do we consider Smith's, et al., estimate of to be generalizable. Our smallest estimate is obtained by assuming the worst possible case of editorial censorship (Appendix B). The estimate is .16 or roughly 25% of the effect size we were led to expect from the Smith et al. (1980) book.

Finally we turn to issues of sampling and generalizability. We found very little overlap of these studies with the MEDLINE data base. We have already pointed out that Smith et al. (1980) overrepresented studies involving academic problems. We also note their overrepresentation of counseling psychologists and educational counselors as opposed to studies more closely tied to medicine (involving clinical psychologists and psychiatrists). We expect smaller effect sizes in the medical setting where emotional problems tend to be longer term and more difficult to resolve. Out of 1759 effect sizes reported in Smith, et al. (1980), 65% involved therapists with a degree in psychology, 22% an education degree, and 13% a degree in medicine with specialization in psychiatry. Yet psychologists and psychiatrists are equally represented among the total population of mental health professionals (Goldenberg, 1983, p.6) - therefore psychiatry has been grossly underrepresented here. We may also draw upon the comparisons between psychotherapy alone versus a control group reported in a separate section of Smith et al. (1980) for studies involving drug

treatments. We expect these studies to reflect the more traditional medical setting and, in support of our position, we find that Smith et al. (1980, p.165) report an average effect size of .30 for these studies.

In conclusion, while overzealous therapists and the news media are quick to generalize the Smith et al. (1980) conclusions to the treatment of the most severe emotional problems, and to all types of settings, our results and those of Smith et al. (1980) for studies involving cases under medical supervision caution against such inferential leaps. Such conclusions must await further analyses of the data that address inferences about populations explicitely.

REFERENCES

Alper, T.G. and Kranzler, G.D. A comparison of the effectiveness of behavioral and client-centered approaches for the behavior problems of elementary school children. Elementary School Guidance and Counseling, 1970, 5, 35-43.

Andrews, W.R. Behavioral and client-centered counseling of high-school underachievers. Journal of Counseling Psychology, 1971, 18, 93-96.

Ashby, J.D., Ford, D.H., Guerney, B.G., and Guerny, L.F. Effects on clients of a reflective and a leading type of psychotherapy. Psychological Monographs, 1957, 71, 1-32.

Beckstrand, P.E. TA as a means of teaching writing in high school. Transactional Analysis Journal, 1973, 3, 161-163.

Carkuff, R.R. and Truax, C.B. Lay mental health counseling. Journal of Counseling Psychology, 1965, 29, 426-431.

Carney, F.J. Evaluation of psychotherapy in a maximum security prison. Seminars in Psychiatry, 1971, 3, 363-375.

Champney, T.F. Adjustment for selection bias in quantitative research synthesis/meta-analysis. American Statistical Association Proceedings of the Social Statistics Section, 1983, in preparation.

Coche, E. and Douglas, A.A. Therapeutic effects of problem solving training and play-reading groups for hospitalized patients. Journal of Psychology, 1977, 33, 820-827.

Cohen, A.C. Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples. Annals of Mathematical Statistics, 1950, 21, 557-569.

Cronbach, L.J. Designing evaluations of educational and social programs. San Francisco: Jossey-Bass, 1982.

Dorfman, E. Personality outcomes of client-centered child therapy. Psychological Monographs: General and Applied, 1958, 72, 1-21.

Eysenck, H.J. An exercise in meta-silliness. American Psychologist, 1978, 33, 517.

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

Glass, G.V., McGaw, B., and Smith, M.L. Meta-analysis in social research. Beverly Hills, Calif.: Sage, 1981.

Goldenberg, H. Contemporary clinical psychology, 2nd Ed. Monterey, Calif.: Brooks/Cole, 1983.

Greenwald, A.G. Consequences of prejudice against the null hypothesis. Psychological Bulletin, 1975, 82, 1-20.

Hedges, L.V. Estimation of effect sizes from a series of independent experiments. Psychological Bulletin, 1982, 92, 490-499.

Hedges, L.V. An analogue to the analysis of variance for effect size data. Journal of Educational Statistics, 1983, in press.

Hedges, L.V. and Olkin, I. Clustering of effect magnitude from independent studies. Psychological Bulletin, 1983, in press.

Jesness, C.F. Comparative effectiveness of behavior modification and transactional analysis programs for delinquents. Journal of Consulting and Clinical Psychology, 1975, 43, 758-779.

Landman, J.T. and Dawes, R.M. Psychotherapy outcome: Smith and Glass' conclusions stand up under scrutiny. American Psychologist, 1982, 37, 504-516.

Lane, D.M. and Dunlap, W.P. Estmating effect size: Bias resulting from the significance criterion in editorial decisions. British Journal of Mathematical and Statistical Psychology, 1978, 31, 107-112.

Rao, C.R. Linear statistical inference and its applications. New York: John Wiley, 1973.

Lazarus, A.A. Behavior rehearsal vs. non-directive therapy vs. advice in effecting behavior change. Behavior Research and Therapy, 1966, 4, 209-212.

Ney, P.G., Palvesky, A.E., and Markeley, J. Relative effectiveness of operant conditioning and play therapy in childhood schizophrenia. Journal of Autism and Childhood Schizophrenia, 1971, 1, 337-349.

Rogers, C.R., Gendlin, E.T., Kiesler, D., and Truax, C.B. The therapeutic relationship and its impact: A study of psychotherapy with schizophrenics. Madison: University of Wisconsin Press, 1967.

Roth, R.M., Mauksch, H.O., and Preiser, K. The non-achievement syndrom, group therapy and achievement change. Personnel and Guidance Journal, 1967, 46, 393-398.

Seeman, J., Barry, E., and Ellinwood, C. Interpersonal assessment of play therapy outcome. Psychotherapy: Theory, Research, and Practice, 1964, 1, 64-66.

Sheldon, W.D. and Landsman, T. An investigation of non-directive group therapy with students in academic difficulty. Journal of Consulting Psychology, 1950, 14, 210-215.

Smith, M.L. and Glass, G.V. Meta-analysis of psychotherapy outcome studies. American Psychologist, 1977, 32, 752-760.

Smith, M.L., Glass, G.V., and Miller, T.I. The benefits of psychotherapy. Baltimore: Johns Hopkins University Press, 1980.

Truax, C.B. Effects of client-centered therapy with schizophrenic patients: Nine years pretherapy and nine years posttherapy hospitalization. Journal of Consulting and Clinical Psychology, 1970, 35, 417-422.

Traux, C.B. and Wittmer, J. Patient non-personal reference during psychotherapy and therapeutic outcome. Journal of Clinical Psychology, 1971, 27, 300-302.

Traux, C.B., Wargo, D.G., and Silber, L.D. Effects of group psychotherapy with high accurate empathy and nonpossessive warmth upon female institutionalized delinquents. Journal of Abnormal Psychology, 1966, 71, 267-274.

Traux, R.A. and Tourney, G. Male homosexuals in group psychotherapy. Disorders of the Nervous System, 1971, 32, 707-711.

Varble, D.L. and Landfield, A.W. Validity of the self-ideal discrepancy as a criterion measure for success in psychotherapy - a replication. Journal of Counseling Psychology, 1969, 16, 150-156.

Appendix A.

General information:

STUDY:

Publication date:

Publication form:

1	2	3	4
journal	book	thesis	unpublished

CLIENTS

Major diagnosis:

neurotic or complex phobic	1
simple phobic	2
psychotic	3
normal	4
character disorder	5
delinquent of felon	6
habituee	7
mixed	8
unknown	9
emotional/somatic complaint	10
handicapped	11
depressive label	12

Mean age to nearest year:

Solicitation of clients:

autonomous presentation	1
presentation in response to advertisement	2
solicited by experimentor	3
committed	4
referred	5

DESIGN

Group assignment of clients:

1	2	3	4	5
random	matching	pretest equation	convenience sample	other nonrandom

Group assignment of therapists:

1	2	3	4	5
random	matching	nonrandom	single therapist	not applicable

Number of comparisons in this study:

Number of this comparison:

Number of outcome measures within this comparison:

Treatment:

placebo	2
psychodynamic	3
client-centered	4
Adlerian	5
gestalt	6
systematic desensitization	7
cognitive/Ellis	8
cognitive/other	9
transactional analysis	10
behavior modification	11
eclectic/dynamic	12
eclectic behavioral	13
reality therapy	14
vocational/personal development counciling	15
cognitive behavioral	16
implosion	17
hypnotherapy	18
other	19

Type of comparison:

1	2	3
control	placebo	treatment

Type of control group:

1	2	3	4	5
no treatment	waiting list	intact group	hospital maintenance	other

Comparison treatment (listed above):

Allegiance of E to therapy compared:

1	2	3
yes	no	unknown

Location of treatment:

School	1
hospital	2
mental health center	3
other clinic	4
other outpatient	5
private	6
other	7
unknown	8
college mental health	9
prison	10
resident m. h. facility	11

Duration of therapy in hours:

Duration of therapy in weeks:

EFFECT SIZE INFORMATION

Sample size for treatment group:

Sample size for comparison group:

Total number in comparison:

Number of weeks posttherapy measure was taken:

Type of outcome measure:

Reactivity of measure:

low	1
high	2

Calculation of effect size:

mean diff/control S.D.	1
MS within	2
MS total minus treatment	3
probit	4
chi square	5
T table	6
mean and P	7
nonparametrics	8
correlations	9
raw data	10
estimates	11
other	12

Source of means

unadjusted post-test	1
covarince adjusted	2
residual gains	3
pre-post differences	4
other	5

Significance of treatment effect:

Treatment group pre-mean:

Treatment group pre-standard deviation:

Treatment group post-mean:

Treatment group post-standard deviation:

Comparison group pre-mean:

Comparison pre-standard deviation:

Comparison post-mean:

Comparison post-standard deviation:

T statistic:

F statistic:

mean square within, residual, or common:

Treatment group percentage improved:

Comparison group percentage improved:

Effect size:

Appendix B

Maximum likelihood estimates of the population effect size are easily obtained through the use of large-sample normal approximations. Hedges (1982, p.492) gives the large-sample normal approximation of the effect size distribution. A variance stabilizing transformation that renders the variance idependent of the mean has been derived by Hedges and Olkin (1983, in press). We use this normal approximation and variance stabilizing transformation earlier in this paper to perform a cluster analysis. The variance stabilizing transformation has been modified by Champney (1983, in preparation) for the case of unequal treatment and control group sample sizes. This transformation is given in the methods section. The variance of this transformation is obtained from the delta theorem given in Rao (1973, p.385):

$$Q_i^2 = t'(\delta_i)^2 \sigma^2(\delta_i)$$

where Q_i^2 is the variance of the transformed variate, $t'(\delta_i)^2$ is the square of the first derivative of the transformation, and $\sigma^2(\delta_i)$ is the variance of δ_i. By substitution we obtain:

$$Q_i^2 = \{1/[2(n_i^E + N_i^C)]\}$$

Note that Q_i^2 does not depend on δ_i. Cohen (1950) has derived the likelihood of a truncated normal variable. The log likelihood, as we have adapted it to the present problem is:

$$\sum_{i=1}^{k} \{-\log[P_i(\delta)] - \log[Q_i/\sqrt{(2\pi)}] - [t(g_i) - t(\delta_i)]^2/2Q_i\}$$

where k is the number of observed studies and Pi(δ) is the area under the normal curve corresponding to observed studies. We use this equation to estimate a common population effect size δ from the distribution with truncation determined by the ∝ level (say p<.05) used as the editorial criterion. To accomplish this we adopt trial values of δ by a grid search method until we obtain a value, $\tilde{\delta}$, that maximizes the likelihood. The method, its justification and rationale, and simulations to examine its small sample behavior are described in detail in Champney (1983, in preparation).

To apply the method to psychotherapy effect sizes we assume a worst possible case situation. All stidues with statistically significant effect sizes are assumed to arise from a truncated normal distribution. Nonsignificant effect sizes are assumed to arise from an untruncated distribution. By adding the likelihoods for these effect sizes we are able to obtain an overall estimate of d. Several ∝ levels are examined to determine what ∝ level results in the smallest estimate of effect size. We also compute $\tilde{\delta}$ assuming that all of the studies arise from an untruncated distribution (i.e. the magnitude of effects is not used as an editorial criterion). The same 18 studies the we include in the cluster analysis are included here.

Our results are given in Figure B-1. With an ∝ level of .05, 8 of the effect sizes are statistically significant and our estimate of δ is .167. It is apparent from Figure B-1 that $\tilde{\delta}$ reaches its smallest possible value for an ∝ level of .08, resulting in 10 significant effect sizes and $\tilde{\delta}$ of .158. Note that as the ∝ level approaches 1.00,

δ approaches the untruncated case. Our maximum likelihood estimate of δ in the untruncated case is .274. This is close to the weighted estimate of .252 that we obtain using Hedges' (1982, p.494) formula without the sample bias correction. With the correction, Hedges formula gives an affect size estimate of .249 as indicated previously in our paper. As we elaborate in the main body of the paper, the homogeneity assumption under which the above effect size estimates are computed is highly unlikely ($p<.000001$).

While the above results serve to illustrate the consequences to meta-analyses of basing editorial decisions on the statistical significance of treatment effects we musr caution against the conclusion that the results of psychotherapy can be summarized by a single effect size estimate. We anticipate that viewing psychotherapy effect size in the population as a random variable itself (a Bayesian view) may be more appropriate. A Bayesian representation of the distribution of population effect sizes is currently under development (Champney, 1983, in preparation). By assuming some α level as a publication criterion we are able to estimate the mean and variance of the effect size distribution and in addition the number of studies that were prevented from publication. Preliminary results suggest that larger mean effect sizes are estimated under these conditions.

Table B-1: Maximum Likelihood Estimates of Effect Size Assuming a Mixture of Studies Including Studies Selected for Publication at Signifiance Level Alpha.

alpha	effect size estimate	# significant
1.00	.274	18
.99	.273	18
.90	.266	17
.50	.231	14
.20	.182	12
.10	.166	10
.09	.162	10
.08	.158	10
.07	.161	9
.05	.167	8
.01	.212	3
.006	.234	2
.003	.251	1
.001	.248	1
.0001	.243	1

```
c*****************************************************************
c
C
C
C                        EFFANOV
C
C                   A FORTRAN/77 PROGRAM
C
C                         FOR
C
C              ANALYSIS OF VARIANCE OF EFFECT SIZES
C
C             COMPILED ON A DEC PDP 11/70 MINICOMPUTER
C              WITH ASCII-STANDARD FORTRAN/77 COMPILER
C
C
C
C                   MATTHEW SCHULZ
C                   MESA PROGRAM
C                   DEPARTMENT OF EDUCATION
C                   UNIVERSITY OF CHICAGO
C                   SEPTEMBER 15, 1982
C
C*****************************************************************
C
C
C   ------------------   HOUSEKEEPING  ------------------------
C
C
      IMPLICIT REAL*8(A-H,O-Z)
      REAL*8 NC,NE
      DIMENSION ARAY(20,20),BRAY(20,20),CRAY(20,20),DRAY(20,20),
     C ERAY(20),KRAY(20)
      OPEN(UNIT=3,NAME='EFFECT.LST',TYPE='NEW') !FOR PRINTOUT
      OPEN(UNIT=1,TYPE='OLD',NAME='EFFECT.DAT') ! CONTAINS E.S., N-CONT, N-EXPER
      OPEN(UNIT=2,TYPE='OLD',NAME='CATEG.DAT') !CONTAINS CATEGORY MEMBERSHIP OF STUDIES
C
C    ---------------- DATA ENTRY AND INITIAL CALCULATIONS --------
C
C  READ IN AVERAGE RELIABILITY OF DEPENDENT VARIABLE MEASURES
C
      R = 1 !RELIABILITY OF OUTCOME MEASURES
C
C  WRITE TABLE HEADINGS FOR PRINTOUT
C
      WRITE(3,1000)
```

```
C
C READ IN DATA AND DO INITIAL CALCULATIONS
C
20     READ(1,*,END=10)G,NC,NE
       READ(2,*) NCAT !CATEGORIES NUMBERED 1,2,...,MAXN
        IF(NCAT.GT.MAXN) MAXN=NCAT !THE NUMBER OF CATEGORIES
       KRAY(NCAT)=KRAY(NCAT)+1  !COUNTS NUMBER OF E.S.'s PER CATEGORY
       K=KRAY(NCAT)
       A1=(NC+NE)/(NC*NE*R)    !ONE COMPONANT OF THE E.S. VARIANCE
       A2=G**2/(2*(NC+NE))  !OTHER COMPONANT OF E.S. VARIANCE
       A=A1+A2  !THE VARIANCE OF AN EFFECT SIZE
       B=1/A  !AS ENTERS THE ANOVA EQUATIONS
       C=G/A  !AS ENTERS THE ANOVA EQUATIONS
       D=G**2/A  !AS ENTERS THE ANOVA EQUATIONS
C
C STORE THESE VALUES FOR EACH E.S.
C
       ARAY(NCAT,K)=A
       BRAY(NCAT,K)=B
       CRAY(NCAT,K)=C
       DRAY(NCAT,K)=D
C
C WRITE IMPORTANT VALUES, AT THIS POINT, TO THE PRINTOUT
C
       WRITE(3,2000)NCAT,NC,NE,G,A,B,C,D
C
C AND REPEAT FOR EACH E.S.
C
       GO TO 20
C
C------------------- WEIGHTED E.S. OF CATEGORY --------
C
C
C WRITE HEADINGS FOR SECOND TABLE OF PRINTOUT
C
10     WRITE(3,3000)
C
C CALCULATE THE WEIGHTED E.S. FOR EACH CATEGORY
C
       DO 1, I=1,MAXN
C
C INITIALIZE NUMERATOR AND DENOMINATOR TO ZERO
C
       CTOT=0
       BTOT=0
C
C THEN TAKE NUMERATOR AND DENOMINATOR SUMMATIONS WITHIN CATEGORIES
C
       DO 2, J=1,KRAY(I)
       CTOT=CTOT+CRAY(I,J)
       BTOT=BTOT+BRAY(I,J)
2      CONTINUE
C
C CALCULATE AND STORE WEIGHTED E.S. OF CATEGORY 'I'
C
       ERAY(I)=(CTOT/BTOT)
```

```
C
1       CONTINUE
C
C------------------------ E.S. ANOVA -------------------
C
C
C BEGIN WITH LOWEST, PROCEED TO HIGHEST CATEGORY
C
        DO 3, I=1,MAXN
C
C INITIALIZE TERMS IN ANOVA EQUATIONS TO ZERO
C
        BTOT=0
        CTOT=0
        DTOT=0
        F=0
C
C CALCULATE SUMS AS ABOVE
C
        DO 4, J=1,KRAY(I)
        BTOT=BTOT+BRAY(I,J)
        CTOT=CTOT+CRAY(I,J)
        DTOT=DTOT+DRAY(I,J)
        F=F+ERAY(I)**2/ARAY(I,J)
4       CONTINUE
C
C WRITE THESE INTERMEDIATE VALUES TO PRINTOUT TABLE 2
C
        WRITE(3,4000) I,BTOT,CTOT,DTOT,ERAY(I),F
C
C CALCULATE SUMS ACCROSS CATEGORIES FOR ANOVA EQUATIONS
C
        HAN=HAN+CTOT
        HAD=HAD+BTOT
        HA1=HA1+DTOT  !THE SUMMED STANDARDIZED E.S.'s
        HB1=HB1+F  !THE SUMMED WEIGHTED, STANDARDIZED E.S.'s
3       CONTINUE
        HT=HA1-(HAN**2/HAD)  !THE TOTAL SUM OF SQUARES
        HB=HB1-(HAN**2/HAD)  !THE BETWEEN GROUPS SUM OF SQUARES
        HW=HT-HB  !THE WITHIN GROUPS SUM OF SQUARES
C
C WRITE FINAL VALUES TO TABLE 3 OF PRINTOUT
C
        WRITE(3,5000)
        WRITE(3,6000) HAN,HAD,HA1,HB1,HT,HB,HW
        CALL EXIT
```

```
C
C---------------------- FORMAT STATEMENTS ---------------
C
1000   FORMAT(1H ,T2,'CATEGORY',T15,'NC',T24,'NE',T32,'G',
     CT41,'A',T50,'B',T59,'C',T68,'D')
2000   FORMAT(1H ,T5,I2,T13,F4.0,T22,F4.0,3X,5(1X,F7.3))
4000   FORMAT(1H ,T5,I2,T13,5(2X,F7.3))
3000   FORMAT(1H0,T5,'CAT',T15,'BTOT',T25,'CTOT',T34,'DTOT',
     CT44,'ERAY(I)',1X,'F')
5000   FORMAT(1H0,T5,'HAN',T15,'HAD',T25,'HA1',T35,'HB1',T45,'HT',
     CT55,'HB',T63,'HW')
6000   FORMAT(1H ,T5,F7.3,T15,F7.3,T25,F7.3,T35,F7.3,T45,F7.3,
     CT55,F6.3,T63,F6.3)
C
C
       END
```

```
C*************************************************************
C
C
C
C                        CLUSTER
C
C                  A FORTRAN/77 PROGRAM
C
C                          FOR
C
C                 GROUPING EFFECT SIZES
C
C  COMPILED ON A DEC PDP 11/70 USING ASCII-STANDARD FORTRAN/77 COMPILER
C
C
C                    MATTHEW SCHULZ
C                    MESA PROGRAM
C                    DEPARTMENT OF EDUCATION
C                    UNIVERSITY OF CHICAGO
C                    APRIL 16, 1983
C
C
C*************************************************************
C
C
C     ------------      HOUSEKEEPING      -----------------
C
C
        IMPLICIT REAL*8(A-H,O-Z)
        REAL*8 NAVGTO,NNTOT,NAVG,NN,NC(50),NE(50)
        DIMENSION CR1(50),CR2(50),EF1(20,20),EF2(20,20),
     * NUM1(20,20),NUM2(20,20),U1(20,20),U2(20,20),
     * U(50),H(50),N1(50),N2(50),GU(50),G(50)
        OPEN(UNIT=1,NAME='EFFECT.DAT',TYPE='OLD')  !CONTAINS E.S., N-CONT, N-EXPER
        OPEN(UNIT=2,NAME='CR1.DAT',TYPE='OLD') !CONTAINS STANDARD NORMAL ORDER STATISTICS
        OPEN(UNIT=3,NAME='CR2.DAT',TYPE='OLD') !CONTAINS BONFERRONI STATISTICS
C
C     ------------  PRELIMINARY CALCULATIONS ------------
C
C READ IN STANDARD NORMAL AND BONFERRONI STATISTICS
C
        DO 14 I=1,100
        READ(2,*,END=15) CR1(I)
        READ(3,*,END=15) CR2(I)
14      CONTINUE
C
C READ IN EFFECT SIZES AND GROUP SIZES FOR EACH STUDY
C
15      DO 1 K=1,1000
        READ(1,*,END=10) G(K),NC(K),NE(K)
C
C     COMPUTE UNBIASED EFFECT SIZE
C
        GU(K)=G(K)*(1-3/(4*(NE(K)+NC(K)-1)))
```

```
C
C   COMPUTE AVERAGE GROUP SIZE IN A STUDY
C
      NAVG=(NE(K)+NC(K))/2
C
C   KEEP RUNNING TOTAL OF SQRT(NAVG) EXCEPT FOR VERY LARGE STUDIES
C
      IF(NAVG.GT.100) GO TO 21
      N=N+1
      NNTOT=NNTOT+SQRT(NAVG)
      NAVGTO=NAVGTO+NAVG  !FOR WEIGHTED CALCULATIONS
C
C   COMPUTE VARIANCE STABILIZING TRANSFORMATION
C
21    A=NC(K)/NE(K)
      DK2=(2*(1+A)**2)/A
      H(K)=GU(K)/SQRT(DK2)
C
C   KEEP RUNNING TOTAL OF TRANFORMATIONS
C
      IF(NAVG.LE.100)HTOT=HTOT+H(K)*NAVG  !FOR WEIGHTED CALCULATIONS
C     HTOT=HTOT+H(K)  !FOR UNWEIGHTED CALCULATIONS
C
C
1     CONTINUE
10    K=K-1  !SETS K TO THE NUMBER OF CASES
C
C
C   COMPUTE AVERAGE STUDY SIZE, LEAVING OUT LARGE STUDIES
C
      NN=(NNTOT/N)**2
      WRITE(5,*) 'AVERAGE STUDY SIZE IS ',NN  !PROGRESS CHECK
C
C   COMPUTE AVERAGE OF V.S. TRANSFORMATIONS
C
      DO 16 I=1,K
      IF((NE(K)+NC(K))/2.LE.100)GO TO 16
      HTOT=HTOT+H(K)*NN
      NAVGTO=NAVGTO+NN
16    CONTINUE
      HBAR=HTOT/NAVGTO
C
C   COMPUTE STANDARDIZED NORMAL DEVIATES
C
      DO 2 I=1,K
      U(I)=2*SQRT(NN)*(H(I)-HBAR)
2     CONTINUE
C
C   WRITE A TABLE OF THE RESULTS UP TO THIS POINT
C
      WRITE(5,2000)
      DO 3 I=1,K
      WRITE(5,2001)I,G(I),GU(I),H(I),U(I),NE(I),NC(I)
3     CONTINUE
```

```
C
C  ---------- DISJOINT CLUSTERING ----------
C
      N=1  !THE ADDRESS OF THE LOWEST SNV IN CURRENT CLUSTER
      J=0  !COUNTER FOR NUMBER OF CLUSTERS
C
C   COMPARE DIFFERENCES BETWEEN S.N.V'S. FROM BOTTOM OF RANGE, UP
C
      DO 4 I=N,K-1
      DIFF=ABS(U(I+1)-U(I))
      IF(DIFF.GT.CR1(K)) GO TO 20
4     CONTINUE
      I=K
C
C   WHEN A SIG. DIFF IS FOUND, STORE CURRENT CLUSTER IN ARRAYS
C
20    J=J+1  !THE CLUSTER NUMBER
      L=0  !COUNTER AND ADDRESS FOR THE NUMBER OF VALUES IN THE CLUSTER
      DO 5 I2=N,I
      L=L+1
      EF1(J,L)=GU(I2)   !THE UNBIASED EFFECT SIZE
      NUM1(J,L)=I2  !THE STUDY NUMBER
      U1(J,L)=U(I2)   !THE STANDARDIZED NORMAL DEVIATE
5     CONTINUE
      N1(J)=L !THE NUMBER OF VALUES IN THE CLUSTER
      IF(I.EQ.K) GO TO 30
      N=I+1  !SO THE NEXT CLUSTER BEGINS WITH NEXT HIGHEST SNV
      GO TO 4
C
C  WRITE A TABLE SHOWING DISJOINT CLUSTERS
C
30    WRITE(5,3000)
      DO 6 I=1,J
      WRITE(5,3001) I,N1(I)
      WRITE(5,3002)
      DO 7 L=1,N1(I)
      WRITE(5,3003)L,NUM1(I,L),EF1(I,L),U1(I,L)
7     CONTINUE
6     CONTINUE
C
C
C   ---------- OVERLAPPING CLUSTERING ----------
C
C
      L=0 !THE NUMBER OF CLUSTERS
      J2=K-1  !THE HIGHEST ALLOWABLE ADDRESS AT BOTTOM OF A CLUSTER
C
C   BEGIN WITH LARGEST AND DECREASE; COMPARE TO LOWEST AND INCREASE
C
      DO 8 I=K,2,-1
      DO 9 J=1,J2
      DIFF=ABS(U(I)-U(J))
      IF(DIFF.LE.CR2(K-J+1)) GO TO 40
9     CONTINUE
      IF(J.EQ.1)GO TO 50
8     CONTINUE
```

```
C
C   STORE CLUSTER IN ARRAYS
C
40      L=L+1 !THE NUMBER OF THE CLUSTER
        N=J   !THE ADDRESS OF THE LOWEST VALUE IN THE CLUSTER
        J2=J-1 !RESETS THE HIGHEST ALLOWABLE ADDRESS FOR NEXT CLUSTER
        N2(L)=I-J+1  !THE NUMBER OF VALUES IN THE CLUSTER
        DO 11 M=1,I-J+1
        NUM2(L,M)=N !THE ADDRESS AND STUDY NUMBER
        EF2(L,M)=GU(N) !THE UNBIASED EFFECT SIZE
        U2(L,M)=U(N)   !THE SNV
        N=N+1
11      CONTINUE
        IF(J.EQ.1)GO  TO 50
        GO TO 8
C
C   WRITE A TABLE OF THE OVERLAPPING CLUSTERS
C
50      WRITE(5,4000)
        DO 12 I=1,L
        WRITE(5,4001) I,N2(I)
        WRITE(5,4002)
        DO 13 M=1,N2(I)
        WRITE(5,4003)M,NUM2(I,M),EF2(I,M),U2(I,M)
13      CONTINUE
12      CONTINUE
        CALL EXIT
C
C
C   ------------   FORMAT STATEMENTS  ------------------
C
2000    FORMAT(1H ,'STUDY',T13,'BIASED',T24,'UNBIASED',T39
     *'H',T51,'U',
     *T60,'N',T67,'N'/1H ,'NUMBER',T14,'E.S.',T26,'E.S.',
     *T35,'TRANSFORM',T47,'STD.N.VAR.',T59,'EXP',T65,'CONT')
2001    FORMAT(1H ,T3,I2,T14,F5.3,T26,F5.3,T37,F5.3,T50,
     *F5.2,T59,F4.0,T67,F4.0)
3000    FORMAT(//1H ,'DISJOINT CLUSTERING'/)
3001    FORMAT(/1H ,T8,'CLUSTER NUMBER ',I2,'  N = ',I2/)
3002    FORMAT(1H ,T16,'MEMBER',T25,'STUDY',T34,'E.S.',
     *T43,'STD.N.VAR.')
3003    FORMAT(1H ,T18,I2,T27,I2,T34,F6.3,T45,F6.3)
4000    FORMAT(//1H ,'OVERLAPPING CLUSTERING'/)
4001    FORMAT(/1H ,T8,'CLUSTER NUMBER ',I2,'  N = ',I2/)
4002    FORMAT(1H ,T16,'MEMBER',T25,'STUDY',T34,'E.S.',
     *T43,'STD.N.VAR.')
4003    FORMAT(1H ,T18,I2,T27,I2,T34,F6.3,T45,F6.3)
        END
```