

Hypnosis as an Adjunct to Cognitive–Behavioral Psychotherapy for Obesity: A Meta-Analytic Reappraisal

David B. Allison and Myles S. Faith
Obesity Research Center
St. Luke's–Roosevelt Hospital Center
Columbia University College of Physicians and Surgeons

I. Kirsch, G. Montgomery, and G. Sapirstein (1995) meta-analyzed 6 weight-loss studies comparing the efficacy of cognitive–behavior therapy (CBT) alone to CBT plus hypnotherapy and concluded that “the addition of hypnosis substantially enhanced treatment outcome” (p. 214). Kirsch reported a mean effect size (expressed as d) of 1.96. After correcting several transcription and computational inaccuracies in the original meta-analysis, these 6 studies yield a smaller mean effect size (.26). Moreover, if 1 questionable study is removed from the analysis, the effect sizes become more homogeneous and the mean (.21) is no longer statistically significant. It is concluded that the addition of hypnosis to CBT for weight loss results in, at most, a small enhancement of treatment outcome.

Kirsch, Montgomery, and Sapirstein (1995) meta-analyzed 18 studies comparing the efficacy of cognitive–behavior therapy (CBT) alone to CBT plus hypnotherapy for a variety of problems. Six of these studies pertained to obesity. They concluded that “the addition of hypnosis substantially enhanced treatment outcome” and that “effects seemed particularly pronounced for treatments of obesity, especially at long-term follow-up, indicating that unlike those in nonhypnotic treatment, clients to whom hypnotic inductions had been administered continued to lose weight after treatment ended” (Kirsch, Montgomery, and Sapirstein, 1995, p. 214). This latter conclusion is the focus of the present article.

Any evidence that there is an efficacious treatment for obesity that results in well-maintained weight loss is cause for enthusiasm, given that the prevalence of obesity is increasing (Kuczmarski, Flegal, Campbell, & Johnson, 1994). However, such evidence merits careful examination, given that long-term weight loss is notoriously difficult to achieve (Brownell & Rodin, 1994) with all but surgical interventions.

Several factors led us to be skeptical of Kirsch et al.'s (1995) conclusion. First, it is inconsistent with the conclusion of several qualitative reviews of this topic (Cochrane, 1992; Mott & Roberts, 1979; Vanderlinden & Vandereycken, 1994; Wadden & Anderton, 1982). Second, the reported weighted mean effect size (standardized mean difference; d) of 1.96 was surprisingly

large, as were two individual effect sizes ($d = 3.65$ and $d = 5.57$). By way of contrast, Abelson reviewed published meta-analyses and found that “it is unusual for this measure [d] to be as big as 1.0, quite rare for it to be as big as 1.4, and extraordinary for it to be as big as 2.0” (1995, p. 89). Third, the distribution of reported effect sizes appeared to be markedly heterogeneous both within and among studies.

Fourth, Kirsch et al. (1995, Figure 1) suggested absolutely no weight regain after hypnosis even at 2-year follow-up. Such perfectly horizontal lines are generally not even found in studies of bariatric surgery. Furthermore, these 2-year follow-up data were based on a single study (Bolocofsky, Spinler, & Coulthard-Morris, 1985). Among the remaining studies, follow-up lengths were limited to 3 and 6 months. Fifth, not all studies appeared to sample clinically obese clients. Specifically, the mean weight of participants in one study was 70 kg (Bolocofsky et al., 1985), which barely qualifies as obesity even among 25-year-old women. It is also noteworthy that 24% of clients in that study dropped out of treatment in the first 9 weeks, an especially high attrition rate for that brief time period. Equally problematic is the fact that participants in Goldstein's (1981) study were allowed to come for additional weekly treatment if desired, which might have confounded this study's effect size estimates. Finally, several unpublished dissertations found that hypnosis was no more effective than cognitive–behavioral interventions in the treatment of obesity. These studies were not included in Kirsch et al.'s meta-analysis, but were reviewed elsewhere (Wadden & Anderton, 1982).

For these reasons, we chose to re-meta-analyze the obesity studies meta-analyzed by Kirsch et al. (1995).

Method

Sample

The sample consisted of the same six studies Kirsch et al. (1995) meta-analyzed. These studies are listed in Table 1. All six are random-

David B. Allison and Myles S. Faith, Obesity Research Center, St. Luke's–Roosevelt Hospital Center, Columbia University College of Physicians and Surgeons.

This study was supported in part by National Institutes of Health Grants P30DK26687, RO1DK40414, and R29DK47256-01A1. We are grateful to Philip Bornstein, Bernard S. Gorman, Louis H. Primavera, David M. Rindskopf, and Lynn Edlen-Nezin for their kind assistance with this article.

Correspondence concerning this article should be addressed to David B. Allison, Obesity Research Center, St. Luke's–Roosevelt Hospital, 1090 Amsterdam Avenue, New York, New York 10025.

Table 1
Comparative Results Using d and Weights Equal to n

Study	Dropout rate (%)	Kirsch's N	Present n	Type of CBT	Kirsch's d	Current d	Low d	High d
Bornstein & Devine (1980)	25	18	18	Covert modeling	0.83	0.07 ^a		
Deyoub & Wilkie (1980)	27	48	35 ^b	Imagery & coping suggestions	-0.17	-0.43 ^c		
Goldstein (1981) [hypnosis with arm levitation]	18	40	40	Self-monitoring + stimulus control + self-reinforcement	5.57	0.95 ^d	0.54	1.37
Goldstein (1981) [hypnosis without arm levitation]	18	40	40	Self-monitoring + stimulus control + self-reinforcement	-0.08	-0.01 ^c	-0.01	-0.02
Wadden & Flaxman (1981)	9	22	20 ^f	Covert modeling	-0.22	-0.21 ^g		
Bolocofsky et al. (1985)	30	109	109	Self-monitoring + goal setting + stimulus control	3.65	0.27 ^h		
Barabasz & Spiegel (1989) [standard suggestions]	NA	30	30	Self-monitoring + goal setting	0.62	0.62		
Barabasz & Spiegel (1989) [individualized suggestions]	NA	29	29	Self-monitoring + goal setting	0.75	0.75		
Weighted mean					1.96	0.28	0.23	0.33
p value for d					<.0001	0.02	0.05	0.01
$\chi^2(7)$ for heterogeneity					152.29	12.89	9.26	18.51
p value for χ^2					<.0001	0.07	0.23	0.01

Note. CBT = cognitive-behavioral treatment.

^a For this comparison the only information offered in the original article was "No . . . differences between groups was found ($p > .10$)" (Bornstein & Devine, 1980, p. 274). Kirsch et al. (1995) apparently obtained the estimated d of .83 by finding the effect yielding a p value of exactly .10. However, it does not seem any more reasonable to assume that a result simply labeled as not significant falls exactly on the cusp of significance than it does to assume it is zero. Thus, we contacted Bornstein, who provided us with the original analysis of variance (ANOVA) tables, and we calculated the effect size from the ANOVA table for weight loss.

^b It appears that Kirsch et al. (1995) used the N that started the trial, not the N that completed the trial. However, the data reported in the original study were only available for participants who finished the trial, so this is the appropriate N to use.

^c We are unable to account for the discrepancy but have carefully checked our calculations, which were based on the individual raw weight loss data provided by Deyoub and Wilkie (1980) in their article. Deyoub and Wilkie (1980) also reported follow-up data. However, this was only for a subset of participants and standard deviations (SD s) or statistics from which they could be calculated were not provided. We, therefore, used only the weight loss data. However, even if these follow-up data were included and the SD s at the earlier time were used as proxies, the average d would not equal -.17.

^d Goldstein (1981) provided a table of means and an ANOVA table but no SD s for weight loss (the dependent variable). Kirsch et al. (1995) also estimated the pooled within-cell SD from the ANOVA table but apparently used a formula designed for a one-way between-groups ANOVA, whereas the data were from a two-way analysis of covariance (ANCOVA) with one within-group factor and one between-group factor. Our effect sizes were obtained by estimating the average within-cell SD by using the mean square errors (MSEs; see Winer et al., 1991, p. 525). However, because Goldstein's analysis used weight at baseline as a covariate, the resulting values should be "unadjusted" for the error variance reducing effects of the covariate; that is, effect sizes derived from ANOVAs and ANCOVAs should not be synthesized because ANCOVA results will probably yield overestimates (Rosenthal, 1991, p. 25). Unfortunately, the correlation of the covariate with the dependent variables was not given. Thus, we calculated plausible upper and lower boundaries around the SD by setting r^2 equal to .90 and .00, respectively. Our "current" estimate of d was based on the average of the d s estimated as the plausible bounds.

There are additional reasons to be cautious regarding the effect sizes obtained from Goldstein (1981). There appears to be a transposition of data in the table of means that Goldstein presented. Other apparent inconsistencies and typographical errors in the results tables suggest the need for skepticism. We contacted Goldstein for more information, but he was unable to provide any. At his suggestion, we also checked the journal for the 4 years after his publication to see if any errata were published but found none.

^e As mentioned earlier, our effect sizes were obtained by using the mean square errors in the equations from Winer et al. (1991, p. 525) and setting plausible bounds on the baseline dependent variable correlation.

^f As mentioned earlier, it appears that Kirsch et al. (1995) used the N that started the trial, not the N that completed the trial.

^g The differences between these two estimates is trivial and is probably due to rounding error.

^h As with Goldstein's (1981) article, this effect size was estimated by calculating the MSEs from the information provided in the article and then solving for the pooled within-cell SD by means of the equations from Winer et al. (1991). Kirsch et al. (1995) used a similar approach but used a formula designed for a one-way between-groups ANOVA, whereas the data were from a two-way ANOVA with one within-groups factor and one between-groups factor.

ized trials. The average dropout rate was 18.2% among those studies reporting dropout rate. Because statistics are based only on those clients who completed treatment, results may overestimate the efficacy of treatment (Allison, Cappelleri, & Carpenter, in press).

Statistical Methods

Kirsch et al. (1995) used d as an effect size metric, where it is the mean difference divided by the (usually pooled) within-group standard

deviation. Weighted analyses were conducted where the weighting factors were sample sizes (N s), per Hunter and Schmidt (1990). In the interests of consistency, we initially conducted our analyses with the same methods. The analysis was then repeated using Hedges and Olkin's (1985) unbiased estimator of d , which we refer to as d_h (note that what is referred to as d here is referred to as g by Hedges & Olkin, and what is referred to as d_h here is referred to as d by Hedges & Olkin). In this case, we also used Hedges and Olkin's weighting scheme, which entails weighting each observation by the inverse of its variance.

Table 2
Results Using d_h and Weights Equal to the Inverse of the Variance

Study	Dropout rate (%)	N	Type of CBT	d_h
Bornstein & Devine (1980)	25	18	Covert modeling	0.07
Deyoub & Wilkie (1980)	27	35	Imagery & coping suggestions	-0.42
Goldstein (1981) [Hypnosis with arm levitation]	18	40	Self-monitoring + Stimulus control + Self-reinforcement	0.93
Goldstein (1981) [Hypnosis without arm levitation]	18	40	Self-monitoring + Stimulus control + Self-reinforcement	-0.01
Wadden & Flaxman (1981)	9	20	Covert modeling	-0.20
Bolocofsky et al. (1985)	30	109	Self-monitoring + goal setting + stimulus control	0.27
Barabasz & Spiegel (1989) [standard suggestions]	NA	30	Self-monitoring + goal setting	0.60
Barabasz & Spiegel (1989) [individualized suggestions]	NA	29	Self-monitoring + goal setting	0.73
All studies				
Weighted mean				0.26
p value for d_h				0.02
$\chi^2(7)$ for heterogeneity				12.32
p value for χ^2				0.09
Without Goldstein (1981)				
Weighted mean				0.21
p value for d_h				0.10
$\chi^2(5)$ for heterogeneity				7.36
p value for χ^2				0.20

Note. CBT = cognitive-behavioral treatment; NA = not applicable.

To be consistent with Kirsch et al.'s approach, we handled multiple effect sizes within studies in the same manner that they did. Specifically, for two studies which had more than one hypnosis condition compared with CBT (i.e., Barabasz & Spiegel, 1989; Goldstein, 1981), a separate d value was calculated for each condition, and no attempt was made to account for the dependency among these observations. For studies involving multiple posttreatment time points, a separate d was calculated for each time point and these d s were averaged. For studies involving multiple dependent variables (e.g., absolute weight and percent overweight) a separate d was calculated for each dependent variable and these d s were averaged.

We conducted tests of heterogeneity using methods described by Hedges and Olkin (1985). When possible, computation of effect sizes and statistical integration was conducted with software from Johnson (1989) and Schwarzer (1988).

Results

Table 1 displays the results of the present analysis compared to those from Kirsch et al. (1995). Where discrepancies between the two analyses occur, we attempt to explain them in the Table 1 footnote. As the table indicates, several discrepancies are present and these appear to be due to computation errors (see Table 1, footnotes a and c through f) and transcription errors (see Table 1, footnotes b and f) in the original meta-analysis.

For one study (Goldstein, 1981), insufficient information was available to definitively calculate an effect size (see Table 1, footnote d). Thus, we calculated the lower bound and a plausible upper bound, included these bounds in Table 1, and recalculated the summary statistics, using these bounds as a form of sensitivity analysis (Greenhouse & Iyengar, 1994). Specifically,

our lower bound estimate assumed an association, r^2 , of zero between the covariate (i.e., baseline weight) and posttest weight, whereas our plausible upper bound assumed an r^2 of .90. We selected a criterion of .90 because body mass index (kilograms per square meters) tends to be a very stable measurement (Heymsfield, Allison, Heshka, & Pierson, 1995) and because Goldstein also reported a "high positive correlation . . . between Ss' weight before and after treatment" (1981, p. 19).

The previous analysis (Kirsch et al., 1995) obtained a mean weighted d of 1.96 ($p < .0001$). Cohen (1988) characterizes d s above 0.8 as "large." However, the results were markedly heterogeneous, $\chi^2(7) = 152.29$ $p < .0001$. The present analysis obtained a mean weighted d of 0.28 ($p = .02$). According to Cohen (1988), a d of 0.20 represents a small effect. Moreover, the results tended to be more homogeneous, $\chi^2(7) = 12.89$, $p = .07$. Overall results were not dramatically different if they were recalculated using the lower and plausible upper bound for the one questionable effect size (see Table 1).

We then repeated the analyses using Hedges and Olkin's (1985) unbiased estimator of d and weights equal to the inverse of each observation's variance. The d_h values were based on the "current" d values from Table 1. Results are displayed in Table 2. The mean weighted d_h was 0.26 ($p = .02$). However, the chi-square test was still suggestive of some heterogeneity, $\chi^2(7) = 12.32$, $p = .09$. The study contributing most to this chi-square was Goldstein (1981), and we had misgivings about the adequacy of effect sizes calculated from the information presented in that article (See Table 1, footnote d). Thus, we eliminated the effect sizes from that study and recalculated the results.

Without this study, the mean weighted d_h was 0.21 ($p = .10$) and the chi-square test of heterogeneity was no longer significant, $\chi^2(7) = 7.36, p = .20$.

Discussion

These results suggest that the previous conclusions regarding hypnosis as an adjunct to CBT for obesity treatment were erroneous. When hypnosis is added to CBT, there appears to be a small effect on average. Moreover, if one questionable study is eliminated, the results are no longer statistically significant.

The present findings are consistent with other quantitative (e.g., Ayyad & Andersen, 1994; Cogan & Rothblum, 1993) and narrative reviews (e.g., Brownell, 1993; Garner & Wooley, 1992; Wilson, 1994) of the broader obesity treatment literature. These reviews suggest only moderately successful outcomes at long-term follow-up. In conjunction with these other articles, the clinical implications of the present meta-analysis are that there is currently no panacea for the treatment of obesity and that hypnosis is no exception.

References

- Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: Erlbaum.
- Allison, D. B., Cappelleri, J. C., & Carpenter, K. M., (in press). Design and analysis in obesity treatment and prevention trials. In S. Dalton (Ed.), *Obesity treatment*. Gaithersburg, MD: Aspen.
- Ayyad, C., & Andersen, T. (1994). A comprehensive literature study of long-term efficacy of dietary treatment of obesity. *International Journal of Obesity*, 18, 78.
- Barabasz, M., & Spiegel, D. (1989). Hypnotizability and weight loss in obese subjects. *International Journal of Obesity*, 8, 335–341.
- Bolocofsky, D. N., Spinler, D., & Coulthard-Morris, L. (1985). Effectiveness of hypnosis as an adjunct to behavioral weight management. *Journal of Clinical Psychology*, 41, 35–41.
- Bornstein, P. H., & Devine, D. A. (1980). Covert modeling-hypnosis in the treatment of obesity. *Psychotherapy: Theory, Research and Practice*, 17, 272–276.
- Brownell, K. D. (1993). Whether obesity should be treated. *Health Psychology*, 12, 339–341.
- Brownell, K. D., & Rodin, J. (1994). The dieting maelstrom: Is it possible and advisable to lose weight. *American Psychologist*, 49, 781–791.
- Cochrane, G. (1992). Hypnosis and weight reduction: Which is the cart and which is the horse? *American Journal of Clinical Hypnosis*, 35, 109–118.
- Cogan, J. C., & Rothblum, E. D. (1993). Outcomes of weight-loss programs. *Genetic, Social, and General Psychology Monographs*, 118, 387–415.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Deyoub, P. L., & Wilkie, R. (1980). Suggestion with and without hypnotic induction in a weight reduction program. *International Journal of Clinical and Experimental Hypnosis*, 28, 333–340.
- Garner, D. M., & Wooley, S. C. (1992). Confronting the failure of behavioral and dietary treatments of obesity. *Clinical Psychology Review*, 11, 729–780.
- Goldstein, Y. (1981). The effect of demonstrating to a subject that she is in a hypnotic trance as a variable in hypnotic interventions with obese women. *International Journal of Clinical and Experimental Hypnosis*, 29, 15–23.
- Greenhouse, J. B., & Iyengar, S. (1994). Sensitivity analysis and diagnostics. In H. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 383–398). New York: Sage.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Heymfield, S. B., Allison, D. B., Heshka, S., & Pierson, R. N., Jr. (1995). Assessment of human body composition. In D. B. Allison (Ed.), *Handbook of assessment methods for eating behaviors and weight-related problems* (pp. 515–560). Thousand Oaks, CA: Sage.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Johnson, B. T. (1989). *DSTAT: Software for the meta-analytic review of research literatures*. Hillsdale, NJ: Erlbaum.
- Kirsch, I., Montgomery, G., & Sapirstein, G. (1995). Hypnosis as an adjunct to cognitive-behavioral psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 63, 214–220.
- Kuczmarski, R. J., Flegal, K. M., Campbell, S. M., & Johnson, C. L. (1994). Increasing prevalence of overweight among U.S. adults. *Journal of the American Medical Association*, 272, 205–211.
- Mott, T., & Roberts, J. (1979). Obesity and hypnosis: A review of the literature. *American Journal of Clinical Hypnosis*, 22, 3–7.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.
- Schwarzer, R. (1988). Meta-analysis programs. *Behavior Research, Methods, Instruments, & Computers*, 20, 338.
- Vanderlinden, J., & Vandereycken, W. (1994). The (limited) possibilities of hypnotherapy in the treatment of obesity. *American Journal of Clinical Hypnosis*, 36, 248–257.
- Wadden, T. A., & Anderton, C. H. (1982). The clinical use of hypnosis. *Psychological Bulletin*, 91, 215–243.
- Wadden, T. A., & Flaxman, J. (1981). Hypnosis and weight loss: A preliminary study. *International Journal of Clinical and Experimental Hypnosis*, 29, 162–173.
- Wilson, G. T. (1994). Behavioral treatment of obesity: Thirty years and counting. *Advances in Behavior, Research, & Therapy*, 16, 31–75.
- Winer, B. J., Brown, D. R., Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.

Received May 31, 1995

Revision received August 14, 1995

Accepted September 4, 1995 ■