Comparison Of Effect Sizes Of Three Group Treatments For Weight Loss

[1]Tianna K. Byom, Ph.D., [2]Marty Sapp, Ed.D.

The current study assessed the efficacy of hypnosis as an adjunct to a standard cognitive-behavioral program for the treatment of obesity. Weight loss was compared among participants who received the standard cognitive-behavioral treatment (CBT), those who received hypnosis in addition to the standard cognitive-behavioral treatment (H-CBT), and a minimal treatment control group focusing on nutrition and exercise information (C). Treatment consisted of eight consecutive weekly group sessions of 1.5 hours and weight was recorded at the beginning of each session. A large effect size (d = .85) was found between H-CBT and C, a moderate effect size (d = .61) was found between CBT and C, and a small effect size (d = .28) was found between H-CBT and CBT. Thus, both the H-CBT group and the CBT group outperformed the minimal treatment group. A small effect size was found between H-CBT and CBT. **(Sleep and Hypnosis 2013;15(1-2):1-10)**

Key words: Hypnosis and weight loss, cognitive-behavioral, effect sizes, confidence intervals

INTRODUCTION

Overweight and obesity rates are increasing at an alarming rate in the United States, doubling in the past 30 years. Overweight and obesity are defined based on body mass index (BMI), which is calculated as weight (kg) divided by height^2 (m). A BMI range of 18.5 to 24.9 is considered healthy. Overweight is defined as a BMI of 25 to 29.9 and obese is defined as a BMI of at least 30. A BMI over 35 is considered morbidly obese. Results from the 2005–2006 National Health and Nutrition Examination Surveys (NHANES) indicate that approximately 32.7% of American adults age 20 and over are overweight, 34.3% are obese, and 5.9% are extremely obese (BMI of 40 or higher). This problem also extends to children and adolescents with 8% to 13% of preschoolers and 13% to 22% of children and adolescents considered overweight, and an additional 31% at risk for becoming overweight (American Heart Association, (1); (11); National Institute of Health, 21; Centers for Disease Control and Prevention (CDC, 9; 11).

[1]Elkhorn, Nebraska USA
[2]Professor, University of Wisconsin–Milwaukee, Department of Educational Psychology, USA

Address reprint requests to: Dr. Marty Sapp, Department of Educational Psychology, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin–Milwaukee, 53211, USA
Tel: (414) 229–6347; Fax: (414) 229–4939;
E–mail: Sapp@uwm.edu

Accepted February 21, 2013

A meta–analysis by Kirsch, Montgomery, and Sapirstein (20) examined 18 studies (1974 to 1993) that compared a cognitive–behavioral treatment with the same treatment with the addition of hypnosis. Presenting problems included pain, insomnia, hypertension, anxiety, obesity, snake phobia, self–concept and athletic performance, duodenal ulcer, and public speaking anxiety. The mean effect size across the studies was .87 standard deviations, indicating that hypnosis significantly increased the efficacy of the cognitive–behavioral treatment. These studies of obesity (8 studies) had the largest effect sizes and were omitted from the calculation of overall effect size in order to give a more conservative estimate. The most conservative estimate was an approximate .5 standard deviations, which suggests that the average client receiving cognitive–behavioral hypnotherapy benefited more than at least 70% of clients receiving cognitive–behavioral treatment alone. Another study rebukes this assertion, however, maintaining that including hypnosis only leads to a small effect size on average (1).

Allison & Faith (1) conducted a meta–analysis on six of the obesity studies (2 studies were not included due to their questionability) reported in Kirsch et al. (20) and found an effect size of .28, which is considered to be small. Kirsch (20) re–conducted his meta–analysis and found an effect size of .98, which is different from the prior two meta–analyses, but still indicates a large effect when hypnosis is combined with cognitive–behavioral treatments. Sapp et al. (29) addressed the conflicting findings by providing confidence intervals that give the upper and lower limits of hypnosis in the treatment of obesity. They found a 95% confidence interval around the population d of

(–.46, .95) for the Allison and Faith study, and (–.04, 1.94) for the Kirsch study. This indicates that both Allison and Faith (1) and Kirsch (19) were correct with their point estimates of effect, but confidence intervals illustrate how their results overlapped. The Allison and Faith study represents the lower limit of effect and the Kirsch study represents the upper limit of effect. Sapp et al. (29) noted that the power level was low for both studies and that additional studies are needed. They also stated that a limitation of the current research regarding using hypnosis to treat obesity is that many of the studies only used female participants and male subjects need to be included in further research.

Purpose of Study

The proposed study was to determine the effect size of hypnosis as an adjunct to a standard cognitive–behavioral program for the treatment of obesity (6). Weight loss was compared among participants who received the standard cognitive–behavioral treatment (CBT), those who received hypnosis in addition to the standard cognitive–behavioral treatment (H–CBT), and a minimal treatment control group focusing on nutrition and exercise information (C). All groups took part in a hypnotic induction prior to treatment

METHODOLOGY

Population

The population is undergraduate and graduate students at large urban university in the Midwest.

Sample

Participants were recruited through flyers posted on–campus, e–mail, and through class recruitment advertising a minimal cost weight loss program. They were informed that they would be participating in an 8–week study comparing the effectiveness of three different treatment groups on weight loss. Recruitment materials included a brief description of the three treatment groups, a description of the initial group hypnosis session that all subjects would participate in, and the minimal risks associated with their participation. These

participants were informed that there was a $20 deposit that would be collected at the initial session and that was fully refundable if they attended every group session. It was stressed that even one missed session would forfeit their deposit. The following were listed as exclusion criteria:

- Involvement in another weight loss program
- Taking weight loss supplements or any medications contraindicated for weight loss
- Has a medical condition that does not allow for moderate–intensity exercise
- Is being treated or has been treated for an eating disorder
- Is not motivated to lose weight

Independent Variables

The independent variables in this study are the three treatment groups. One group received a standard cognitive–behavioral treatment for weight loss adapted from the LEARN Program for Weight Management (7). The second group received the identical cognitive–behavioral treatment with the addition of hypnosis. Subjects experienced a hypnotic induction and were given suggestions (imagery, ego–strengthening) for weight loss. In addition, participants were given suggests to decrease calories, and to eat foods high in fiber so that they would have a sense of fullness. It was suggested to participants to loss one to three pounds a week. Participants were also given suggestions to keep a diet journal. That is, they were given suggestions to monitor their caloric intake. Participants were given suggestions to eat slowly so that their brains could register fullness. Moreover, it was suggested to participants to eat at prescribe times in a certain room so they could better monitor caloric intake. Participants were also given suggestion to exercise two to three times per week for at least 30 minutes. Self–hypnosis suggestions were also given to participants. Finally, they received relapse prevention suggestions. For example, it was suggested to participants that they would make mistakes, and they were given suggestions to recover and start over and not be discouraged. The third group received a minimal treatment of nutrition and exercise information.

Dependent Variable

The dependent variable was body weight and was recorded by measurement on a standard weight scale.

Procedure

Thirty–seven individuals (32 female, 5 male) attended an initial session and informed consent was obtained. The Waterloo–Stanford Group Scale of Hypnotic Susceptibility, Form C (WSGC), was administered by a graduate student trained in hypnosis in order to assess for hypnotic suggestibility. Hypnosis was explained to the group prior to administering the WSGC in order to clear up any misconceptions about hypnosis, faulty treatment expectancies, and possible side effects of hypnosis. After administering the WSGC, participants completed the corresponding form and the Inner Subjective Experience Ratings Scale (in order to assess for nonvolitional responses). Participants were taken individually to the corner of the room, weighed on a standard scale, and the $20 deposit was collected. They were also engaged in conversation in order to verify that they were no longer in a hypnotic state. Participants were weighed until the same weight was displayed twice (which was usually obtained by the second reading). Because the results of the ANOVA did not suggest weight category differences, subjects were randomly assigned to one of the three treatment groups. Immediately before the first group session, 3 individuals (1 male from hypnosis group, 2 females from minimal treatment group) reported that they would not be able to participate in the study. Two individuals stated that conflicting schedules prevented them from participating in the study

and the other individual cited personal reasons for not participating. The group sessions began with 30 females and 4 males.

Cognitive-Behavioral Group

Participants in the cognitive–behavioral group received a standard cognitive–behavioral treatment based on the LEARN Program for Weight Management (Brownell, 7). The group met for 8 weekly sessions of 1.5 hours starting Tuesday, March 25, 2008. The group started with 12 female participants, but only 10 participants completed the study. Reasons cited for leaving the study included lack of time and a schedule that conflicted with the group meeting time.

Hypnosis Group

The hypnosis group was identical to the cognitive–behavioral group but included a hypnotic induction at the end of each session that gave suggestions based on the content of the session and provided imagery (i.e. imagining being at ideal body weight and associated feelings, imagining making healthy food choices, etc.) and ego–strengthening. The participants were instructed to practice self–hypnosis at least once a day and the therapist checked in to make sure participants had been using self–hypnosis. The group started with 10 females and 2 males, but only 6 females and 1 male completed the study. Conflicting schedule and need for increased study time were reasons given for prematurely leaving the group. Three others left the group with no explanation.

Minimal Treatment Group

The minimal treatment group met for the same number of sessions and time, but they only received basic information about nutrition, exercise and environment. The rest of the time was devoted to group discussion about progress. The minimal treatment group gained the basic tools they needed in order to lose weight, but they did not learn specific techniques like the other groups and they were responsible for how they interpreted the information. They came up with their own goals and did not receive the charts and forms like the other groups. The group started with 8 females and 2 males, but only 4 females and 1 male completed the study. Reasons for leaving the group included a desire to be in the hypnosis group, lack of motivation, and unknown reasons.

Measures

Waterloo–Stanford Group Scale of Hypnotic Susceptibility, Form C (WSGC) – The WSGC was developed as a substitute for the individually administered Stanford Hypnotic Susceptibility Scale, Form C (SHSS:C) and contains 12 items. The response booklet requires participants to estimate whether or not they objectively responded to the 12 suggestions. Items are scored with either a 0 (did not objectively experience) or a 1 (objectively experienced), so scores can range from 0 to 12.

Bowers (5; 6) provided normative data on the WSGC and reported internal consistency at.80, which is slightly less than the .85 reported by Hilgard (15) for the SHSS:C. The WSGC and the SHSS:C were found to correlate .85 with each other and the WSGC had an internal consistency of .81 and the SHSS:C had an internal consistency of .84, illustrating that the two scales correlate as high as the reliabilities of the scales will allow.

Inner Subjective Experience Ratings Scale (ISER) – Developed in order to assess the subjective experiences of subjects to the 12 hypnotic suggestions. Participants rate each suggestion on a 5–point Likert–type scale, with anchors specific to the content of each suggestion. Ratings are summed to produce a total subjective scale score ranging from 12 to 60. Normative data on the subjective scale report an internal consistency of .89 and correlations between the WSGC behavioral and experiential scores were significant, which suggests that the subjective scale is a reliable

and valid measure of experiential responses to hypnotic suggestion (5).

Weight – Measured on a standard scale.

RESULTS

Initial Weight

A total of 34 participants began the study. Prior to treatment, participants were randomly assigned to a treatment group and a one–way ANOVA was conducted to determine if significant differences in weight existed and necessitated stratification of groups. Initial weight was entered as the dependent variable and the treatment group was the independent variable. The ANOVA did not indicate a significant difference between the groups, $F(2,34) = .799$, $p = .459$. Therefore, participants were kept in the original groups.

A one–way ANOVA was run with treatment group as the independent variable and percentage of weight loss as the dependent variable. Results of the ANOVA indicate that there are no statistically significant differences between the three treatment groups, $F(2,22) = 1.161$, $p = .334$.

Planned comparisons illustrate that the H–CBT group and CBT group did not differ significantly from the C group, $t(19) = 1.466$, $p = .159$. A second planned comparison tested if the H–CBT group loss more weight than the CBT group. The H–CBT group did not differ significantly from the CBT group, $t(19) = .541$, $p = .595$.

Power Analyses

Statistical significance refers to the likelihood that the relationship between variables or the difference between means is due to chance. The p value is the probability that a test statistic of at least the same size would have arisen by chance, even if there was actually no difference between the two populations. Not finding a statistically significant result, however, does not mean that there is no difference between the groups. The p value is heavily influenced by both the size of the effect and the size of the sample (23).

When a study has a very low sample size, power is greatly diminished. The power of a study refers to the probability that a given test will find a statistically significant difference when such a difference indeed exists. As power decreases, the chances of making a Type II error (failing to detect a difference when one exists) increases. A target value for statistical power is often set at .80 (80%) so that four times out of five, a false null hypothesis will be correctly rejected (12).

With a total n of 22, there is reason to believe that the present study suffers from low power. A one–way ANOVA with percentage of weight loss as the dependent variable and treatment group as the independent variable revealed a power of .22 or a 22% chance of the study correctly detecting a statistically significant difference. In other words, there is a 78% chance that a Type II error was committed.

Effect Sizes

Effect size is a way to quantify the difference between groups and is not affected by sample size (28; 16; 25). There are many different ways to measure effect size but in general, effect size is calculated as the standardized difference between two means (Cohen's d and Hedges' g) or as the correlation between the independent variable category and the individual scores on the dependent variable, which is called the effect size correlation (24; 14). In addition to its utility in quantifying the effectiveness of a particular intervention, effect size also allows for comparisons across studies. Unfortunately, the reporting of effect size is scarce outside of meta–analysis (18; 28; 31, 32; 33; 28).

Overall Effect Size

The overall effect size was calculated by a formula suggested by Stevens (1999) and is expressed in terms of the F–statistic.

The overall effect size was determined to be .32 and illustrates the impact of the three treatments on weight loss. Cohen (8) classifies an f of .1 as a small effect size, .25 as a medium effect, and anything above .4 as a large effect size. An f of .32 would therefore be considered by Cohen (8) to indicate a moderate effect size.

While it is useful to know the overall effect size, it is also important to examine the effect of each of the treatments. Cohen's d and Hedges' g (unbiased version) were used because they were the measures of effect size used in the meta–analyses of Kirsch (19; 20) and Allison and Faith (1) and their calculations would allow for comparisons to be made. Cohen's d is calculated by subtracting the mean of the control group from the mean of the treatment group and dividing that by the pooled standard deviation.

H-CBT and CBT

Using the means and standard deviations of the H–CBT and CBT groups, a d of .28 was found. Cohen (8) suggested a d effect size of .20 as small, .50 as medium, and .80 is considered a large d effect size. Therefore, there is a small effect size between the H–CBT and CBT groups. A 95% confidence interval around the population of d is (–.68, 1.24). The confidence interval provides the same kind of information as a significance test. Since the confidence interval includes zero, this is the same as saying that the result is not statistically significant.

The analysis was repeated using Hedges and Olkin's (14) correction method and a value of .266 was found. This is nearly identical to Cohen's d value. A 95% confidence interval around the corrected value is (–.70, 1.23).

H-CBT and C

Comparing the H–CBT and C groups resulted in a d of .85, which is a large effect size. A 95% confidence interval around the population of d is (–.34, 2.04), indicating that the result is not statistically significant.

The corrected value of d was found to be .78, which still signifies a large effect size. The 95% confidence interval around the corrected value is (–.4, 1.96).

CBT and C

After comparing the CBT and C groups, a d of .61 was obtained. This is considered by Cohen (1977) to be a moderate effect size. The 95% confidence interval around the population of d is (–.49, 1.71).

The correction formula produced a value of .57 and is still considered to be a moderate effect size. The 95% confidence interval around this value is (–.52, 1.66).

Reliability

In order to assess internal consistency, Cronbach's alpha coefficient was calculated for both the Waterloo–Stanford Group C Scale (WSGC) and the Inner Subjective Experience Ratings Scale (ISER). The results of the reliability analysis provide a Cronbach's alpha of .622 for the WSGC. A 95% confidence interval for the correlation coefficient is (.338, .818).

The Inner Subjective Experience Ratings Scale was found to have a Cronbach's alpha of .74. A 95% confidence interval around the correlation coefficient is (.545, .875).

Correlational Analysis

A bivariate correlation was calculated to assess the relationship between the WSGC and ISER. The correlation analysis revealed a positive correlation between the WSGC and ISER, r =

Table 1. ANOVA to test for significant differences between groups for My Group Facilitator Was Skilled and Knowledgeable.

		Sum of Squares	df	Mean Square	F	Sig.
My group facilitator was skilled and knowledgeable.	Between Groups	6.155	2	3.077	8.009	.003
	Within Groups	7.300	19	.384		
	Total	13.455	21			

.541, n = 22, p = .009.

Participants were asked to rate their level of agreement with the statement, My group facilitator was skilled and knowledgeable, on a scale that ranged from Strongly Disagree (1) to Strongly Agree (6). A statistically significant between group difference was found at F(2,22) = 8.01, p = .003. Group means were 6.00 for H–CBT, 5.70 for CBT, and 4.60 for C.

Participants ranked their concurrence with the statement, Going to the group sessions was pointless – it was a waste of my time, from 0 (No, not at all true) to 7 (Yes, very much true). This item intended to measure their perceived importance of the group. A significant result was obtained at F(2,22) = 7.33, p = .004. Means of the groups were 1.14 for H–CBT, .40 for CBT, and 4.00 for C.

DISCUSSION

Wampold et al. (34) reported that meta–analyses to establish absolute efficacy (that psychotherapy is better than no treatment) yield "a convenient and historic benchmark value of .80" (p. 714). The conclusion is that counseling and psychotherapy are remarkably efficacious. However, Wampold et al. (34) found that when psychotherapies were compared, no significant differences were found between the types of treatment. The effect size for the comparison of therapies was found to be between .00 and .21 (with .21 being a small effect size) and suggests that while treatment is usually significantly better than no treatment, significant differences may not exist among treatments. This supports the idea that it is the common factors among bona fide treatments that lead to treatment outcomes and not the specific factors.

Since only 22 participants completed the study, the study lacked sufficient power to be able to detect a statistically significant difference, even if one existed. Therefore, effect sizes between each of the three groups were calculated. The APA Task Force on Statistical Inference (1999) stressed that effect sizes should "always" be reported along with p values (Wilkinson & Task Force on Statistical Inference, p. 599). The fifth edition of the APA (3) Publication Manual of the American Psychological Association also added, "For the reader to fully understand the importance of your findings, it is almost always necessary to include some index of effect size or strength of relationship in your results section" (pp. 25–26). The APA manual (3) also noted that the "failure to report effect sizes" is a "defect in the design and reporting of research" (p. 5).

The statistician Karl Pearson began to popularize the correlation coefficient around 1896 and in 1962 Jacob Cohen used the letter d to signify the standardized difference between means (17). Sapp (26) reported that there are over 40 different effect size measures and that effect sizes are important because they allow for meta–analytic thinking or a quantitative way of synthesizing the literature within an area of hypnosis. Sapp (27) noted that while there are many ways to calculate effect size, two are commonly used within meta–analysis, the correlation and the d effect size. One measure of effect size is the r effect size measure, which is the correlation between an independent variable and dependent variable. Using the Pearson product–moment correlation coefficient r, effect size ranges from –1 to +1 with 0 representing no effect and –1 and +1 being the maximum effect (either negative or positive). Rosenthal (22) defined an r value of .1 as a small effect size, .3 as a medium effect size, and .5 as a large effect size, but emphasized that these are just rules of thumb and effect sizes must be evaluated within a substantive area. Cohen's d, on the other hand, focuses on the magnitude of the difference between or among group means rather than the strength of the association and Cohen (8) defined d as the difference between two means divided by the pooled standard deviation of both groups. Cohen reported a d of .2 as a small effect size, .5 as a medium effect size, and .8 as a large effect size. Thompson (31) stated, "A very important implication of the realization that there are dozens of effect size

statistics is that authors must explicitly tell readers what effect sizes they are reporting, so that the effects can be properly interpreted and compared apples–to–apples across studies!" (p. 424).

In addition to reporting effect sizes, the APA Task Force on Statistical Inference strongly encouraged the reporting of confidence intervals (CIs) and stressed the importance of interpreting the results of a study through comparison with related results in prior research (p. 599). The APA (3) Publication Manual of the American Psychological Association also suggested that confidence intervals represent "in general, the best reporting strategy. The use of confidence intervals is therefore strongly recommended" (p. 22). Although APA emphasizes the importance of using confidence intervals, empirical studies confirm that confidence intervals are virtually never reported in the social sciences (10). Thompson (31) noted that it is plausible that many researchers do not completely understand statistical methods that are rarely reported in the literature and that they infrequently use in their own work.

As an illustration, many people who use CIs misinterpret a 95% confidence interval as indicating that they can be 95% confident that their confidence interval captures the estimated population parameter. In reality, Thompson (31) noted that "what computing 95% CIs for a statistic means is that, if we drew infinitely many random samples from the population, exactly 95% of the CIs would capture the parameter, and exactly 5% would not" (p. 427). Confidence intervals provide more information than a point estimate, which is a single numerical value that is used as an estimation of the sample statistic. Confidence intervals also give an indication of the precision of the estimates by examining the width of the intervals. A narrow interval indicates a more precise estimate. CIs also assist in meta–analytic thinking by allowing researchers to examine effect sizes and confidence intervals across research studies. This can be helpful in formulating study expectations and research design, and it also allows for retrospective interpretation through comparison with effect sizes from prior research.

When a confidence interval includes zero, some researchers erroneously interpret this in the same way as a statistical significance test (13). Thompson (37) reported that the most informative use of confidence intervals does not evaluate whether a CI includes zero, rather compares the CI to CIs across multiple studies. The benefit of using confidence intervals is that using CIs across studies will eventually reveal the correct population value, even if the initial estimate was completely wrong (Schmidt, 1996). Allison and Faith (1) and Kirsch (19) conducted meta–analyses to assess the effect of adding hypnosis to cognitive–behavioral therapy and produced conflicting results. Sapp et al. (29) put confidence intervals around the effect sizes and found a 95% confidence interval around the population effect size of (–.46, .95) for the Allison and Faith study (1) and (–.04, 1.94) for the Kirsch study (19). This illustrates that the two studies have overlapping confidence intervals, with one study representing the lower limit of effect and the other representing the upper limit of effect. The 95% confidence interval around the effect size between H–CBT and CBT in the current study is (–.68, 1.24). This indicates that the confidence interval of the current study overlaps with the confidence intervals in the meta–analyses conducted by Allison and Faith (1) and Kirsch (19).

It is important to note that the results of the study by Bolocofsky et al. (4) found no significant differences in weight between participants receiving a behavioral management program and participants receiving the same program with the addition of hypnosis immediately after treatment. However, at an 8 month and 2–year follow–up, participants in the hypnosis group continued to lose weight while participants in the behavioral management group without hypnosis did not. Thus, hypnosis did not appear to have a significant effect on weight loss until follow–up. This suggests that the full benefits of hypnosis may not be realized immediately after treatment.

Another limitation of the current study is the lack of follow–up data. The end of the treatment groups coincided with the end of spring semester, and the researcher also moved out of state soon after treatment ended. As a result, valuable follow–up data was not able to be collected. Bolocofsky et al. (4) found that the effects of hypnosis on weight loss were only apparent at an 8 month and 2–year follow–up, suggesting that the major benefits of hypnosis take place long after treatment is over. Since a small effect size of .28 (95% CI of –.68 to 1.24) was already found between the H–CBT and CBT groups, it is quite possible that the H–CBT group would either continue to lose weight at a higher rate than the CBT group or perhaps they would be more effective at maintaining their weight loss. Given that no follow–up data was collected, however, this hypothesis cannot be tested. It is critical for further endeavors to examine treatment effects at later dates.

The most significant limitation of this study is likely its small sample size. Due to the small sample size, the current study lacked the statistical power to detect a difference even if one existed. The power of a statistical test is defined as its probability of rejecting the null hypothesis when it is false or the probability of making a correct decision. While Stevens (30) suggests a power level of .70 is adequate and .90 is excellent, the current study has an overall power level of .22. Power is dependent on many factors including the alpha level, effect size, sample size, the statistical test used, the research design, and the link between the treatment and the dependent variable. Power is heavily influenced by sample size and when the sample group size is small, adequate power can only be achieved if the effect size is extremely large. Stevens (30) noted that researchers who are not sensitive to the issues surrounding power and rely on null hypothesis testing may interpret nonsignificant results as indicating that the treatments involved made no difference. It may very well be that the treatments did make a difference but the study lacked the power to detect the difference.

The current study expanded on the current literature and examined the specific aspects of hypnosis that may contribute to weight loss. For example, participants found the suggestions to be slightly more effective than the imagery and it appears that hypnosis was most effective in increasing feelings of self–control. Results of this study may be used to design hypnotic treatment protocol that, when combined with cognitive–behavioral treatment, may assist individuals in losing more weight than with cognitive–behavioral treatment alone.

REFERENCES

1. *Allison, D. B., & Faith, M. S. (1996). Hypnosis as an adjunct to cognitive-behavioral psychotherapy for obesity: A meta-analytic reappraisal. Journal of Consulting and Clinical Psychology, 64, 513-516.*

2. *American Heart Association. (2005). Heart disease and stroke statistics – 2005 update. Dallas, TX: American Heart Association.*

3. *American Psychological Association. (2001). Publication manual of the American Psychological Association (5th ed.). Washington, DC: Author.*

4. *Bolocofsky, D. N., Coulthard-Morris, L, & Spinler, D. (1984). Prediction of successful weight management from personality and demographic data. Psychological Reports, 55, 795-802.*

5. *Bowers, K. S. (1993). The Waterloo-Stanford Group C (WSGC) Scale of Hypnotic Susceptibility: Normative and comparative data. International Journal of Clinical and Experimental Hypnosis, 41(1), 35-46.*

6. *Bowers, K. S. (1998). Waterloo-Stanford Group Scale of Hypnotic Susceptibility, Form C: Manual and response booklet. International Journal of Clinical and Experimental Hypnosis, 46(3), 250-268.*

7. *Brownell, K. D. (2004). The LEARN program for weight management (10th ed.). Dallas: American Health Publishing Company.*

8. *Cohen, J. (1977). Statistical power analysis for the behavioral sciences. New York: Academic Press.*

9. *Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2005-2006, http://www.cdc.gov/nchs/data/hestat/overweight/overweight_adult.htm.*

10. *Finch, S., Cumming, G., & Thomason, N. (2001). Reporting of statistical inference in the Journal of Applied Psychology: Little evidence of reform. Educational and Psychological Measurement, 61, 181-210.*

11. *Flegal, K. M., Carroll, M. D., Ogden, C. L., & Johnson, C. L. (2002). Prevalence and trends in obesity among US adults, 1999-2000. The Journal of the American Medical Association, 288, 1723-1727.*

12. *Fox, N. J., & Mathers, N. J. (1997). Empowering research: Statistical power in general practice research. Family Practice, 14(4), 324-329.*

13. *Hagen, R. L. (1997). In praise of the null hypothesis statistical test. American Psychologist, 52, 15-24.*

14. *Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. Orlando: Academic Press.*

15. *Hilgard, E. R. (1965). Hypnotic susceptibility. New York: Harcourt, Brace & World.*

16. *Hoyle, R. H. (1999). Statistical strategies for small sample research. Thousand Oaks, CA: Sage.*

17. *Huberty, C. J. (2002). A history of effect size indices. Educational and Psychological Measurement, 62(2), 227-240.*

18. *Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., Kowalchuk, R. K., Lowman, L. L., Petoskey, M. D., Keselman, J. C., & Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. Review of Educational Research, 68(3), 350-386.*

19. *Kirsch, I. (1996). Hypnotic enhancement of cognitive-behavioral weight loss treatments – Another meta-reanalysis. Journal of Consulting and Clinical Psychology, 64, 517-519.*

20. *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.*

21. *National Institute of Health (1998). Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults: The Evidence Report, NIH Publication No. 98-4083.*

22. *Rosenthal, R. (1984). Meta-analytic procedures for social research. Beverly Hills, CA: Sage.*

23. *Rosenthal, R., & Rosnow, R. L. (1991). Essentials of behavioral research: Methods and Data Analysis (2nd ed.). New York: McGraw-Hill.*

24. *Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes and counternulls on other people's published data: General procedures for research consumers. Psychological Methods, 1, 331-340.*

25. *Rosnow, R. L., & Rosenthal, R. (2009). Effect sizes: Why, when, and how to use them. Journal of Psychology, 217(1), 6-14.*

26. *Sapp, M. (2004a). Cognitive-behavioral theories of counseling: Traditional and nontraditional approaches. Springfield, IL: Charles C. Thomas Publisher.*

27. *Sapp, M. (2004b). Confidence intervals within hypnosis research. Sleep and Hypnosis, 6(4), 169-176.*

28. *Sapp, M. (2008). The effect sizes r and d in hypnosis research. European Journal of Clinical Hypnosis, 8(1), 41-59.*

29. *Sapp, M., Obiakor, F., Scholze, S., & Gregas, A. (2007). Confidence intervals and hypnosis in the treatment of obesity. Australian Journal of Clinical Hypnotherapy and Hypnosis.*

30. *Stevens, J. (1999). Intermediate statistics: A modern approach (2nd ed.). Mahwah, NJ:Lawrence Erlbaum Associates.*

31. *Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. Psychology in the Schools, 44(5), 423-432.*

32. *Thompson, B. (2002a). "Statistical," "Practical," and "Clinical": How many kinds of significance do counselors need to consider? Journal of Counseling & Development, 80, 64-71.*

33. *Thompson, B. (2002b). What future quantitative social science research could look like: Confidence intervals for effect sizes. Educational Researcher, 31(3), 25-32.*

34. *Wampold, B. E. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "All must have prizes". Psychological Bulletin, 122(3), 203-215.*