Contemporary Hypnosis
Contemp. Hypnosis 25(2): 94–99 (2008)
Published online 5 June 2008 in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/ch.355

PUBLICATION BIAS IN STUDIES ON THE EFFICACY OF HYPNOSIS AS A THERAPEUTIC TOOL

Morten Moshagen and Jochen Musch

Institute for Experimental Psychology, University of Düsseldorf

Abstract

In an extensive review of more than 400 hypnotic treatment outcome studies, Flammer and Bongartz (2003) presented meta-analytic evidence supporting the efficacy of hypnosis as a therapeutic tool. Meta-analyses, however, are prone to the problem of selective publication of studies reporting positive outcomes. In the present investigation, we therefore employed a variety of methods to test for the presence of publication bias in the data analysed by Flammer and Bongartz (2003). The results suggest that publication bias may have contributed to the effect size estimate by about one third. However, our analysis also shows that the efficacy of hypnosis is of a substantive nature, and may not be explained on the basis of publication bias alone. Copyright © 2008 British Society of Experimental & Clinical Hypnosis. Published by John Wiley & Sons, Ltd.

Key words: hypnotherapy, meta-analysis, publication bias

Introduction

Flammer and Bongartz (2003) have recently analysed an extensive body of research on the efficacy of hypnosis as a therapeutic tool. They based their analysis on a set of 57 randomized controlled clinical trials; in these studies, hypnosis was used either to support medical interventions, or to treat disorders that can be coded according to ICD-10 criteria (World Health Organization, 1992). The meta-analytic review of these hypnotic treatment outcome studies resulted in a pooled effect size estimate of $r = 0.27$, indicating a significant effect of medium size. Meta-analytic reviews of the efficacy of therapeutic interventions, however, may lead to erroneous conclusions if there is evidence for a selective reporting of positive outcomes. For example, a 'file-drawer-problem' (Rosenthal, 1979) occurs when only significant outcomes are published, whereas studies with non-significant outcomes, or outcomes that are contrary to expectations, remain unpublished. Given that such unpublished studies are usually difficult to locate, meta-analytic results may be biased towards an overly optimistic estimate of the magnitude of a treatment effect.

To address the problem of a potential publication bias, Flammer and Bongartz (2003) computed a failsafe N statistic which expresses the number of unpublished studies averaging a zero effect size that would be necessary to nullify their meta-analytic mean effect size estimate (Rosenthal, 1979). The failsafe N, however, suffers from several shortcomings. First, it ignores the sample size of omitted studies, equating studies with a sample of $N = 10$ to those with a sample of $N = 10,000$ (Begg and Berlin, 1988); second, the assumption that the omitted studies average a zero effect has to be questioned (Gleser and Olkin, 1996; Sutton, Song, Gilbody and Abrams, 2000); and finally, the lack of an underlying statistical model accompanied with the lack of a clear-cut statistical criterion limits its utility as a statistical test (Orwin, 1983). For these and other reasons, the failsafe N is neither an informative, nor a sufficient statistic to rule out publication bias (Becker, 2005). Therefore, the purpose of the present investigation was to re-analyse Flammer and Bongartz's (2003) data using methods that are better suited to test for a selective reporting of positive outcomes.

Method

Among several procedures that have been proposed to test for the presence of publication bias, those based on the funnel plot are most widely used. A funnel plot is a scatter plot of the effect sizes of all studies against a measure of their precision (e.g. the standard error of the effect size estimate). Due to a lower degree of random variation, studies with higher precision (i.e. smaller standard errors of the effect size estimate) will show less dispersion around the mean effect size than those with less precision. In the absence of any bias, the funnel plot is shaped like a funnel. If there is a selective inclusion of studies showing positive outcomes, an asymmetry in the funnel diagram arises due to a gap caused by the absence of negative outcomes. Although intuitively appealing, the usefulness of funnel plots to graphically test for the presence of publication bias is limited because of the dependence on a researcher's subjective judgment (Terrin, Schmid and Lau, 2005). Formal tests of funnel plot asymmetry exploit the fact that under conditions of publication bias, a statistically reliable association between study effect size and study precision emerges. Based on this reasoning and implementing a direct statistical analogue of the visual inspection of a funnel plot, the rank-correlation method (Begg and Mazumdar, 1994) utilizes Kendall's rank correlation to examine the association between the standardized effect sizes and their variances. A statistically significant positive correlation between these measures reflects a trend towards bigger effect sizes in studies with smaller samples, and is regarded as indicative of publication bias.

Because of its superior statistical power (Kromrey and Rendina-Gobioff, 2006), we also employed the alternative Egger regression method (Egger, Davey Smith, Schneider and Minder, 1997) in addition to the rank-correlation test. In the Egger regression method, the standard normal deviate (effect size standardized by its standard error) is regressed on its precision (defined as the inverse of the standard error). Because the standard error of effect size estimates is largely determined by sample size, the precision of smaller studies will be low, as will be the standard normal deviate. Hence, the slope of the regression equation indicates the direction and magnitude of the effect, while the intercept provides a measure of the degree of asymmetry in the funnel plot, with larger deviations from zero indicating a larger degree of asymmetry and thus, a larger bias.

Going beyond an assessment of the degree of asymmetry, the nonparametric trim-and-fill method proposed by Duval and Tweedie (2000a, 2000b) aims at estimating the number and outcome of studies that are presumably missing in a meta-analysis as a result of publication bias. Assuming that in the absence of any bias effect sizes are symmetrically distributed around the 'true' effect size, the basic idea of the trim-and-fill method is to estimate the number of 'missing' studies based on the asymmetry of the observed distribution. After augmenting the data set with the presumably missing studies, it is possible to compute a weighted mean effect size adjusted for bias. The putative number of missing studies is estimated in an iterative manner: the outlying asymmetrical part

Published by John Wiley & Sons, Ltd
DOI: 10.1002/ch

of the funnel plot consisting of unusually large effect sizes is temporarily removed ('trimmed'), and the weighted mean effect size is recalculated omitting these trimmed studies. Then, again using the full set of studies, studies that are asymmetrical with respect to the new estimate of the mean are trimmed. The number of studies that need to be trimmed according to this procedure stabilizes after a few iterations. The resulting dataset is then augmented by adding mirror-image counterparts of the trimmed studies to compensate for the missing data. Finally, a new weighted mean effect size adjusted for bias is computed based on this symmetrically augmented dataset, which now includes the temporarily trimmed, the symmetrically added, and all remaining studies.

The aforementioned methods were applied in order to test for the presence of publication bias in the 57 randomized controlled clinical trials included in Flammer and Bongartz's (2003) meta-analysis of the therapeutic efficacy of hypnosis. Fisher Z-transformed correlation coefficients served as a common measure of effect size. Analyses were conducted using version 2.2 of the Comprehensive Meta Analysis software (Borenstein, Hedges, Higgins and Rothstein, 2005).

Results

Figure 1 shows that the funnel plot exhibits signs of asymmetry. This visual impression receives support from the rank-correlation method, which yields a significant estimate of $\tau = 0.21$ ($p < 0.05$).

Employing the linear regression test confirms this analysis and results in a significant intercept of 1.70 ($p < 0.01$). Moreover, employing the trim-and-fill procedure leads to the conclusion that 17 studies seem to be missing in the original data set of 57 studies.

Figure 2 shows the funnel plot after imputing these missing studies, which results in a set of 57 + 17 = 74 studies. The adjusted weighted mean effect size for the augmented data set is $r = 0.18$ ($p < 0.01$) using a fixed effects model, and $r = 0.19$ ($p < 0.01$) using a random effects model.

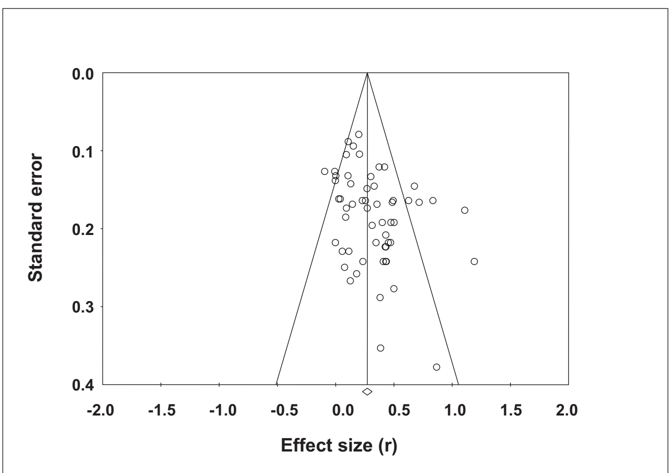

Figure 1. Funnel plot of the original data set.

Published by John Wiley & Sons, Ltd
DOI: 10.1002/ch

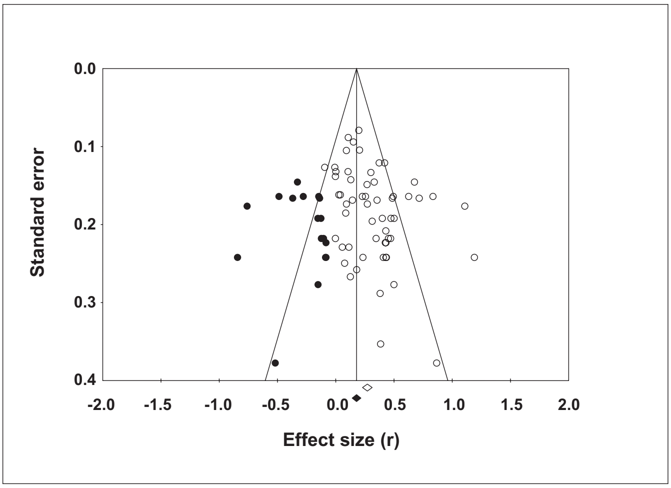

Figure 2. Funnel plot of the augmented data set.
Note: Open circles indicate original data; filled circles indicate studies that were imputed according to the fill-and-trim procedure. Diamonds indicate effect size estimates that are uncorrected (open) or corrected (filled) for publication bias.

Discussion

We considered the possibility that in their meta-analysis of the efficacy of hypnosis as a therapeutic tool, Flammer and Bongartz (2003) overestimated the mean treatment effect due to the selective publication of studies with positive outcomes. Each of the analyses we conducted indeed indicated that publication bias is present in their data set. First, inspection of the funnel plot revealed notable signs of asymmetry, indicating a dearth of medium-sized and small studies with negative outcomes, which is contrary to what would have to be expected assuming normally distributed effect sizes unaffected by selective reporting. The rank correlation method (Begg and Mazumdar, 1994) yielded a statistically significant estimate of $\tau = 0.21$; smaller studies were more likely to report positive treatment outcomes than larger studies. The linear-regression method (Egger et al., 1997) confirmed this finding by showing a significant intercept indicating an association between study outcome and study precision. Finally, the trim-and-fill method suggested that 17 studies had been omitted in the original data set. Imputing the omitted studies led to reduced mean effect size estimates of $r = 0.18$ (fixed-effects model) and $r = 0.19$ (random-effects model), respectively; the magnitude of the original unadjusted fixed and random effect size estimates ($r = 0.27$ and $r = 0.31$, respectively) was thus reduced by about one third.

Finally, a word of caution is necessary in interpreting the results of the present study. Although the selective publication of studies with positive outcomes will manifest itself in an asymmetric funnel plot, funnel plot asymmetry may also arise as a result of true population heterogeneity (Sterne, Gavaghan and Egger, 2000). In case of true between-study heterogeneity causing smaller studies to yield systematically weaker effects than larger studies (which might occur, for example, when a particular intervention is more effective for rare disorders than for common disorders, because studies on rare disorders tend to be smaller than studies on common disorders), the trim-and-fill method may

Published by John Wiley & Sons, Ltd
DOI: 10.1002/ch

underestimate the true treatment effect (Terrin, Schmid, Lau and Olkin, 2003). Unfortunately, to date there is no means to distinguish between possible causes for funnel plot asymmetry.

This limitation notwithstanding, all of the tests we conducted consistently suggested the presence of publication bias in the set of 57 randomized controlled clinical trials analysed by Flammer and Bongartz (2003). However, our results also show that there is a statistically significant treatment effect even after controlling for publication bias using the trim-and-fill-method. We therefore conclude that in spite of a tendency to selectively report positive outcomes, the efficacy of hypnosis seems to be substantial, and its apparent effect may not be explained on the basis of publication bias alone. To avoid erroneous conclusions, however, we recommend that future meta-analyses on the efficacy of therapeutic interventions should routinely employ tests to detect a possible publication bias.

References

Becker BJ (2005) Failsafe N or the file-drawer number. In: HR Rothstein, AJ Sutton, M Borenstein (eds) Publication Bias in Meta-analysis. New York: Wiley, 111–26.

Begg CB, Berlin JA (1988) Publication bias: a problem in interpreting medical data. Journal of the Royal Statistical Society: Series A 151: 419–63. DOI: 10.2307/2982993.

Begg CB, Mazumdar M (1994) Operating characteristics of a rank correlation test for publication bias. Biometrics 50: 1088–101. DOI: 10.2307/2533446.

Borenstein M, Hedges L, Higgins J, Rothstein H (2005) Comprehensive Meta-analysis Version 2. Englewood Cliffs, NJ: Biostat.

Duval S, Tweedie R (2000a) A nonparametric 'trim and fill' method of accounting for publication bias in meta-analysis. Journal of the American Statistical Association 95: 89–98. DOI: 10.2307/2669529.

Duval S, Tweedie R (2000b) Trim and fill: A simple funnel-plot based method of testing and adjusting for publication bias in meta-analysis. Biometrics 56: 455–63. DOI: 10.1111/j.0006-341X.2000.00455.x.

Egger M, Davey Smith G, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. British Medical Journal 315: 629–34.

Flammer E, Bongartz W (2003) On the efficacy of hypnosis: a meta-analytic study. Contemporary Hypnosis 20: 179–97. DOI: 10.1002/ch.277.

Gleser LJ, Olkin I (1996) Models for estimating the number of unpublished studies. Statistics in Medicine 15: 2493–507. DOI: 10.1002/(SICI)1097-0258(19961215)15:23<2493::AID-SIM381>3.0.CO;2-C.

Kromrey JD, Rendina-Gobioff G (2006) On knowing what we don't know: an empirical comparison of methods to detect publication bias in meta-analysis. Educational and Psychological Measurement 66: 357–73. DOI: 10.1177/0013164405278585.

Orwin RG (1983) A fail-safe N for effect size in meta-analysis. Journal of Educational Statistics 8: 157–9. DOI: 10.2307/1164923.

Rosenthal R (1979). The "file drawer problem" and tolerance for null results. Psychological Bulletin 86: 638–41. DOI: 10.1037/0033-2909.86.3.638

Sterne JAC, Gavaghan D, Egger M (2000) Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. Journal of Clinical Epidemiology 53: 1119–29. DOI: 10.1016/S0895-4356(00)00242-0.

Sutton AJ, Song F, Gilbody SM, Abrams KR (2000) Modeling publication bias in meta-analysis: a review. Statistical Methods in Medical Research 9: 421–5. DOI: 10.1191/096228000701555244.

Terrin N, Schmid CH, Lau J (2005) In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. Journal of Clinical Epidemiology 58: 894–901. DOI: 10.1016/j.jclinepi.2005.01.006.

Published by John Wiley & Sons, Ltd
DOI: 10.1002/ch

Terrin N, Schmid CH, Lau J, Olkin I (2003) Adjusting for publication bias in the presence of heterogenity. Statistics in Medicine 22: 2113–26. DOI: 10.1002/sim.1461.

World Health Organization (1992) The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines. Geneva: World Health Organization.

Address for correspondence:
Morten Moshagen, Jochen Musch
Institute for Experimental Psychology
University of Düsseldorf
Universitätsstr. 1, 40225 Düsseldorf, Germany
Tel: +49-211-81-13494/+49-211-81-11524
Fax: +49-211-81-11753
Email: {morten.moshagen,Jochen.musch}@uni-duesseldorf.de

Published by John Wiley & Sons, Ltd
DOI: 10.1002/ch