Psychotherapy Research

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tpsr20

Therapist effects in the National Institute of Mental Health Treatment of Depression Collaborative Research Program and other psychotherapy studies

Paul Crits-Christoph a & Robert Gallop b
a University of Pennsylvania
b West Chester University

Available online: 22 Feb 2007

To cite this article: Paul Crits-Christoph & Robert Gallop (2006): Therapist effects in the National Institute of Mental Health Treatment of Depression Collaborative Research Program and other psychotherapy studies, Psychotherapy Research, 16:02, 178-181

To link to this article: http://dx.doi.org/10.1080/10503300500265025

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Therapist effects in the National Institute of Mental Health Treatment of Depression Collaborative Research Program and other psychotherapy studies

PAUL CRITS-CHRISTOPH¹ & ROBERT GALLOP²

¹University of Pennsylvania and ²West Chester University

(Received 16 March 2005; revised 8 April 2005; accepted 11 April 2005)

We appreciate this opportunity to react to the two studies in this issue of Psychotherapy Research (Elkin, Falconnier, Martinovich, & Mahoney; Wampold) that address therapist effects on outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program (TDCRP; Elkin et al., 1989). The fact that one of these articles (Elkin et al., 2005) reports no reliable therapist outcome differences in the TDCRP and the other (Wampold et al., 2005) reports significant therapist effects, and suggests that therapist effects in the TDCRP are crucial for understanding a Treatment × Initial Severity interaction, indicates that important questions remain about how best to analyze and interpret differences between therapists in the context of a randomized clinical trial testing treatment effects.

Elkin et al. (2005) cogently address a number of misconceptions and methodological pitfalls that have arisen in regard to therapist outcome effects. Our goal in the current commentary is to briefly expand on some of the central issues discussed by Elkin et al. (2005) and to touch on some additional points not covered. In particular, we first address the topic of the Wampold (2005) article, namely the existence of therapist differences in the TDCRP and the implication of such differences for the previously reported (Elkin et al., 1995) Treatment × Severity interaction in the TDCRP. We then discuss the larger question raised by Elkin et al. (2005) regarding the extent of therapist differences in other psychotherapy studies.

Therapist effects in the TDCRP

Why does Wampold (2005) report significant therapist effects in the TDCRP, but Elkin et al. (2005) do not? There are several differences in the hierarchical linear modeling (HLM) approaches in these two articles. For one, Wampold analyzes endpoint scores, whereas Elkin et al. analyze all data collected (e.g., baseline, Week 4, Week 8, Week 12, Week 16). Wampold's analysis was, therefore, a two-level HLM (patient nested within therapist) and Elkin et al.'s was a three-level HLM, including time, patient, and therapist. In addition, Wampold analyzes a completer sample, which presumably consists of Week 16 (termination) scores only. An intention-to-treat (ITT) sample is also analyzed by Wampold, but the author does not directly state how this was done. Given that the intention-to-treat sample consists of all randomized patients (with the exception of one therapist who had one client eliminated), we can infer that a last-observation-carried-forward (LOCF) method was used to impute endpoint scores. LOCF uses the last nonmissing score, regardless of whether it occurred at Week 4, 8, 12, or 16, as the final endpoint score.

There are serious problems with both LOCF analyses and analyses of completer samples (Gibbons et al., 1993; Houck et al., 2004; Lavori, 1992). For the LOCF analysis, the most notable problem is the impact of time. The LOCF analysis disregards the time effect and treats each endpoint as if it was obtained at the same point in time and assumes that patients would have no change beyond that endpoint. The completer analysis does not ignore time, because all patients are assessed at the same point in time (Week 16 in the TDCRP), but the completer analysis requires that all missing values occur completely at random. The "completely at random" assumption means that the probability of an observation being missing does not depend on observed or unobserved measurements. This assumption is more restrictive than the "missing at random" assumption of HLM models (Houck et al., 2004). "Missing at random"
means that missing data do not depend on unobserved measurements but can depend on observed measurements.

These potential biases introduced in such analyses are part of the justification for using longitudinal data analytic methods that include all data collected at multiple time points. Because substantially more data are used in the longitudinal models analyzed by Elkin et al. (2005), such analyses would be expected to have greater statistical power and yield more accurate estimates of change than static analyses of endpoint scores as well as be relatively free of the biases that can occur in LOCF and completer analyses (Diggle, Liang, & Zeger, 1994). With greater statistical power and more accurate estimation of change over time, Elkin et al. (2005) find no significant therapist effects, with effect sizes close to zero on primary measures of depression.

The possibility of various biases in Wampold’s (2005) ITT –LOCF and completer analyses is also suggested by the inconsistencies across these analyses and the various models and outcome measures presented. For example, for the Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960), the percentage of variance resulting from therapists is 8.6 times greater in the completer model with therapist slopes fixed compared with the ITT model with therapist slopes fixed (0.069 vs. 0.008, respectively).

Wampold’s (2005) attempt to illustrate that Elkin et al.’s (1995) reported Treatment × Initial Severity interaction on the HRSD is no longer noteworthy once therapist is included as a random effect. We do not find this analysis compelling. Wampold (2005) begins with a nonsignificant interaction effect (p = .08) when therapist is not included in the model and only the two psychotherapies are examined. A new statistical model including a nonsignificant (p = .37) therapist term as a random effect reduces the p value for the Treatment × Severity interaction to .22. However, the model with therapist as a random effect has reduced statistical power for testing the Treatment × Severity interaction (the degrees of freedom are now based on the number of therapists, not the number of patients), and this reduced power, along with some “noise” added by the nonsignificant therapist effect, may explain the movement of the p value from a nonsignificant .08 to a nonsignificant .22.

We also wish to provide a clarification of Wampold’s (2005) statement that “therapists are more important than treatments.” What the data suggest is that differences between therapists were more important than differences between treatments. This is not to say that treatments were not important. Both treatments may simply be equally efficacious. An assessment of the importance of treatments must be made by means of a comparison to control groups, not by comparing active treatments.

The magnitude of therapist effects

How large are therapist effects in the psychotherapy research literature? Wampold (2005) stated that “the research that has been conducted to estimate therapist effects has shown that a modest to large proportion of the variability in outcome is due to the therapist (Blatt, Sanislow, Zuroff, & Pilkonis, 1996; Crits-Christoph et al., 1991; Luborsky, McLellan, Diguer, Woody, & Seligman, 1997; Project MATCH Research Group, 1998).” However, as Elkin et al. (2005) note, it is necessary to look a bit more closely at this cited evidence. Although Crits-Christoph et al. (1991) reported a modest effect (8.6% of outcome variance was due to therapist across studies), Luborsky et al. (1997) did not report an effect size for therapist, the Project MATCH (Project MATCH Research Group, 1998) effect was due to an outlier (although the outlier differed in different analyses), and Blatt et al. (1996) found no significant differences between therapists in the TDCRP. Thus, the cited evidence does not support the claim of consistent moderate to large therapist differences.

There are some studies in the literature, including some of those in the Crits-Christoph et al. (1991) meta-analysis, that appear to show relatively larger therapist differences in outcome. However, as Elkin et al. (2005) point out, methodological factors (e.g., small sample sizes, use of therapist measures only) may account for some of these apparent larger effects. To this we add that the selection of studies in the Crits-Christoph et al. (1991) meta-analysis, and other reviews that have emphasized the existence of therapist effects, may have been biased toward inclusion of studies that had therapist differences. An examination of the extent of therapist effects using a truly representative review of the literature has never been performed.

In fact, many studies now published do not appear to demonstrate reliable differences between therapists. These results are overlooked by those who search for the existence of therapist effects. In most cases, an article on the topic of therapist differences is published only when such differences are found (e.g., Huppert et al., 2001). For other randomized clinical trials, the lack of therapist effects is typically mentioned in a single sentence concerned with preliminary analysis (e.g., Borkovec, Newman, Pincus, & Lytle, 2002; Crits-Christoph
et al., 1999; Fals-Stewart & O’Farrell, 2003; Stanley et al., 2003; Taylor et al., 2003). Until a comprehensive meta-analysis is conducted, we suggest that if therapist effects are to be discussed, writers should cite some of the large number of nonsignificant differences in the literature rather than only citing a few significant ones, so that a balanced view is given.

However, as Elkin et al. (2005) point out, the question of the impact of the therapist on outcome is better evaluated in the context of naturalistic studies rather than manual-based clinical trials. The large-scale study by Okiishi, Lambert, Nielsen, and Ogles (2003) is the best estimate to date on the magnitude of therapist differences in a naturalistic setting. These authors again emphasized differences between therapists, focusing on those with particularly good outcomes and those with particularly bad outcomes. The overall magnitude of therapist effects was not directly reported by Okiishi et al. (2003) but can be estimated from the parameters given in their Table 3. We calculate that 4.1% of the outcome variance (slopes) was attributed to therapist differences (using the numbers in Table 2: 0.057/[0.057 + 1.338] = 0.041). This is between a small (i.e., 1%) and moderate (i.e., 9%) effect, using Cohen’s (1988) definitions. However, as Okiishi et al. (2003) point out, this effect may underestimate true therapist differences given nonrandom assignment of patients to therapists.

In summary, our best estimate of therapist differences in outcome from both large-scale multicenter clinical trials like the TDCRP (Elkin et al., 2005), Project MATCH (Project MATCH Research Group, 1998), and the NIDA Cocaine Collaborative Study (Cris-t Christoph et al., 1999) is that such effects are typically small; our best estimate from large-scale naturalistic studies (Okiishi et al., 2003) is that such effects are between small and moderate. Nevertheless, research should continue to investigate the conditions under which therapist differences might be apparent in naturalistic settings and routinely conduct preliminary analyses, checking on the presence of therapist effects when analyzing data from efficacy studies.

Acknowledgements

The preparation of this article was funded in part by National Institute on Drug Abuse Grants R01-DA018935 and R21-DA016002.

References


Downloaded by [Dr Jeremy Safran] at 11:54 11 December 2011