Psychotherapy Research

Rejoinder to commentaries by Stephen Soldz and Paul Crits-Christoph on therapist effects

Irene Elkin a, Lydia Falconnier b, Zoran Martinovich c & Colleen Mahoney a

a University of Chicago
b University of Illinois at Chicago
c Northwestern University

Available online: 22 Feb 2007
We are in complete agreement with Soldz that multilevel modeling is very complex and that many options have to be weighed and decisions made as to the exact nature of the models to be used. We have tried, to the best of our ability, in modeling therapist effects in the National Institute of Mental Health Treatment of Depression Collaborative Research Program (TDCRP), to think through these issues very carefully and to get expert statistical consultation when necessary. Crits-Christoph and Soldz both point out that a major decision that had to be made involved the number of levels used in the analyses. Soldz also mentioned the way in which baseline levels are addressed. We would like to further describe the rationale for the choices that we made in regard to both of these issues.

We chose to use a three-level model, with time at the first level, so that we could include data throughout the entire treatment period. As we have previously written, we believe it is important to include data between the first and last time points “since these data can provide a much better picture of how individuals change over time, rather than snapshot images of the first and last time points” (Gibbons et al., 1993, p. 741). In addition, change rates based on multiple times are more reliable than rates based on only two time points, and this increased reliability enhances statistical power. A third advantage of using multilevel growth curve models is that they allow you to use all data for all patients over the course of treatment without the need to use extrapolations or imputations to account for missing data.

Although missing data, most frequently a result of patient attrition, will always be a problem in outcome analyses, the assumptions that are made in regard to this problem when modeling treatment effects are important. Our judgment is that the model form assumptions and estimation procedures of growth curve models are more tenable than the assumptions in endpoint analyses, in which missing data are dealt with by carrying the last data point forward, which effectively assumes no change after a variable endpoint. Our growth model strategy does not carry data forward but rather assumes a continuous, regular change pattern. Estimated individual time trend lines are based on available data for each individual, augmented by information from data for all other individuals in the sample. The hierarchical linear modeling estimation algorithm thus allows cases with relatively more reliable assessments (i.e., more assessments over broader ranges of time) to have greater weight in estimating model parameters. Because we believe that the assumptions of our strategy are more tenable and make more efficient use of the data, we agree with Crits-Christoph’s assessment and are more inclined to trust our growth model findings for the total sample.

By anchoring (i.e., having each patient’s change trajectory pass through his or her baseline score on the outcome variable), we can summarize each patient’s fitted trajectory by a single number: rate of change. Thus, as Soldz has pointed out, we eliminate one source of variance in our model. We view anchoring the intercept as analogous to analysis of covariance (ANCOVA) procedures commonly used in pre–post outcome analyses (i.e., including pretest as a covariate and measuring adjusted outcomes on the posttest variable). In ANCOVA models, adjusted change estimates tend to be more reliable than unadjusted change estimates. Similarly, by anchoring the intercept and eliminating the intercept source of random variation, slope estimate reliability is enhanced, and this improves statistical power. Given the difficulty involved in acquiring large sample sizes, both our anchoring strategy and...
more common ANCOVA approaches are, in part, motivated by power considerations.

We are clearly in agreement with both Crits-Christoph and Soldz about limitations resulting from sample size (of both therapists and patients per therapist) in the TDCRP and in the belief that the future in this area will lie in large data bases (collected, e.g. in managed-care settings). We join Soldz in urging that those involved in these efforts ensure the inclusion of relevant patient and therapist characteristics, so that questions regarding therapist effects and other issues of importance in psychotherapy research can be adequately addressed.

Reference