Process Transforms Inputs to Determine Outcomes: Therapists Are Responsible for Managing Process

Merton S. Krause, Northwestern University
Wolfgang Lutz, University of Trier

The variance in outcomes for psychotherapy patients is not partitionable into components that are independent contributions of treatments, therapists, and patients. If these inputs did not influence one another over the course of psychotherapy, they could be independent and so have additive main-effects or interaction-effects on outcomes. But that is impossible because they do influence one another and therapists are responsible for actively managing the psychotherapy process by repeatedly adjusting these inputs toward optimally influencing one another. The consequent interdependence of these inputs within the therapy process needs to be reflected in the design and analysis of psychotherapy outcome studies, as it presently is not, if we are to learn who is adequate for treating whom, how, and why so.

Key words: **disaggregation, input types, nonrecursiveness, outcomes, process, therapist types, variance partitioning. [Clin Psychol Sci Prac 16: 73–81, 2009]**

We can partition the dependent variable outcome variance of psychotherapeutic treatments in randomized clinical trials (RCTs) *as if* the effects of treatment types, psychotherapist types, and patient types could be disentangled, but it is important to appreciate that they cannot be, for the following three reasons.[1]

First, the types of psychotherapeutic treatment can be validly evaluated for their effectiveness only according to how they are performed by given samples of therapists and how they are engaged by each therapist's given sample of patients. Therefore, types of treatment cannot reasonably be evaluated for effectiveness simply as if in their outcome-relevant regards they were independent of these therapists and patients.

Second, individual therapists and types of therapists can be evaluated for their effectiveness only for what they actually are like and do in their work with their patients, and this must in part be influenced by the types of treatment that are incorporated and emphasized in their work and by the nature of their patients and how these patients participate in the therapy process. Therefore, therapists cannot reasonably be evaluated for effectiveness independently of the particular mixtures of and relative emphases on the types of treatment they employ and of the particular types of patients they treat, which are matters that they are professionally responsible for but are only somewhat under their control.

Third, a patient's responsiveness to treatment depends in part on the actual treatment and relational context provided by her or his particular therapist. Therefore, a patient cannot be evaluated for her or his contribution to the outcome of treatment independently of the mixture of and differential emphasis on the treatment types provided and of the other influences of the therapist on the therapy process.

All of this ought to be quite apparent to anyone who is familiar with the nature of psychotherapy. And it should be taken seriously in the design and analysis of psychotherapy outcome studies. (The role of types of treatment circumstances, such as inpatient or outpatient and voluntary or involuntary, will be ignored here in the interest of simplicity of presentation and because the

Address correspondence to Merton S. Krause, Department of Psychology, Northwestern University, Evanston, IL 60208. E-mail: msk514@msn.com.

reader can easily see that it must parallel the entangled roles of treatment, therapist, and patient types.)

ORTHOGONAL PARTITIONING OF OUTCOME VARIANCE

If there were a mutual causal independence of the independent-variables or factors in a fully crossed or full-factorial RCT design, then the dependent-variable variance could be partitioned into three orthogonal main-effect components: the between-treatments component, the between-therapists component, and the between-patients component (see, e.g., Cox, 1958, pp. 23−47). The presently supposed way to realize such a design in RCTs is to have two or more (supposedly) uniformly fixed types of treatment, the therapists assigned at random to these treatments so that the treatments may receive stochastically equivalent lots of therapists (in terms such as of aptness at providing the treatment), and the patients assigned at random to therapists so that the therapists may receive stochastically equivalent[2] caseloads (in terms such as of responsiveness to the treatment). If all of this were achieved in such a doubly nested design (i.e., therapists nested within treatments and patients nested within therapists, which is also referred to as a three-level hierarchical design), then independent main-effect components of patient outcome variance due to differences in initial input types of treatment, to differences in initial input type mixes of therapists, and to differences in initial input type mixes of patients would be distinguishable. Also independent treatment type by therapist type, therapist type by patient type, treatment type by patient type, and treatment type by therapist type by patient type interaction-effect components of patient-outcome variance (from what was left after the main effects were partitioned or partialled out) would be distinguishable.

This is a kind of mathematical ideal, but one that does not suit the realities of psychotherapy. It does not, because to properly partition *outcome* variance into independent or orthogonal components associated with specific treatment, therapist, and patient initial inputs, each of these kinds of inputs must in fact be unaffected by and in this sense independent of the other kinds of inputs. The simplest way to achieve this in a case is to have each input variable maintained at a fixed level throughout the course of that case. In this way we could logically consider that *these specific initial inputs* account for specific

orthogonal components of *outcome* variance. But this is impossible in psychotherapy outcome studies, as shall be argued next.

THE CAUSAL ENTANGLEMENT OF TREATMENTS, THERAPISTS, AND PATIENTS

Causal entanglement (Elkin, 1999) means that outcome-relevant treatment, therapist, and patient input variables or types causally influence each other, which is the problem of nonrecursiveness (see, e.g., Berry, 1984, pp. 7−18). The appropriateness of the analysis of variance (ANOVA) for estimating the relative efficacy of different treatment types, therapist types, or patient types in an RCT rests (most simply) on the assumption that each independent variable (or contrast, such as between an experimental treatment E and a control treatment C) that can account for some component of dependent variable variance in that RCT is an absolutely (or mutually independently stochastically) stable influence on this variance. If this assumption is true, then the influence on outcomes of each controlled independent variable (or contrast) in that RCT is independent of the influence of the other controlled variables in the study. This means that the contrast of an experimental treatment type (E) versus some control condition (C), that is, the mean difference in outcomes between E and C, is not biased by differences in the therapist types (such as being sufficiently apt at or focused in treating) or their mixes or by differences in the patient types (such as being motivated for or responsive to treatment) or their mixes in these two comparison groups.

However, this independence or orthogonality assumption cannot, in general, reasonably be assumed to hold, because the independent variables or input types actually operative in psychotherapy logically must tend to causally influence one another in the process of psychotherapy. Furthermore, the dependent variables also must tend to influence the independent variables, because what therapists do next in the course of an episode of psychotherapy should and does clinically somewhat depend upon what has already happened to the patient in terms of the outcome variables. And if this independence assumption does not hold well, then ANOVA is an inappropriate form of data analysis for this (doubly nested or three-level) RCT design, and also inappropriate are the hierarchical regression analyses that are now quite commonly used to orthogonally partition outcome variance

by successively partialling out components of variance (see, e.g., Cohen & Cohen, 1983). So psychotherapy cannot reasonably be studied as if it were a system of strictly independently contributing causes whose effects on patient outcomes independently add up and do not causally affect *each other*, that is, as if it were a recursive system as ANOVA logically requires (see, e.g., Berry, 1984, pp. 11–15). Because it is not a recursive system, the total between-treatments and between-therapist-within-treatments components of variance from this usual patients-nested-in-therapists-nested-in-treatments RCT design cannot fairly be interpreted as indicating the extent of the different treatment types' or therapist types' or patient types' *independent* contributions to the variance in the outcomes of psychotherapy.

This nonrecursiveness is obvious in the very nature of psychotherapy. Professional therapists are not simply uniform conveyors of treatment types to patients, but in various manners and to various extents they must try to adapt these treatment types to themselves and to their particular patients and must try to adapt themselves and these patients to these treatment types. And they rationally must do so responsively to their patients' changes in (what the particular therapist at least implicitly takes to be) the relevant process or outcome variables. What competent therapist does not try to optimize the translation of an assigned or chosen treatment type (e.g., how here and now to effectively reinforce constructive behaviors, credibly communicate clinically helpful information, present and time and work through interpretations, convey unconditional positive regard, etc.) and of his or her initial way of relating (e.g., with what warmth, openness, concern, patient-role induction efforts, etc.) according to how she finds she can sustain doing so and according to what the impacts of various versions of these seem to be on each patient.[3] Similarly, patients are not simply uniform recipients of treatment types and of therapists' ministrations, but in various manners and to various extents they must (rationally or not) adjust what they make of and how they deal with therapists and treatment types and adapt themselves to these therapists and treatments, and this must (at least sometimes) be in terms of what each patient (probably implicitly) takes to be the relevant process and outcome variables.[4] What motivated patient does not try to optimally adjust to (or escape) what is offered her or him, according to how helpful (or

intolerable) it seems. This is where the nonrecursiveness necessarily occurs in the processual system of psychotherapy, so no *fixed* and therefore independent therapist or patient influence upon outcome ought be presumed to be added to some *fixed* and therefore independent treatment influence upon outcome.

There are obviously potential if not explicitly manifest dynamisms here, that is, nonrecursiveness, quite unlike the recursiveness (i.e., the *absence* of causal feedback loops of independent or dependent variables causally influencing other independent variables) that logically must exclusively prevail for appropriate applications of ANOVA.[5] Simply equating the mean *initial* values of possibly outcome-relevant patient variables across therapists and of possibly outcome-relevant therapist variables across treatments, either exactly by individual selection and matching or on average by random assignment of patients and therapists (where either of these methods of control *is* actually feasible and successful),[6] could not make a causal system that is nonrecursive be recursive. The complex causal exigencies of the psychotherapeutic situation are likely quite various across patients, episodes of treatment with the same patient, and even sessions within episodes, so the appropriate nonrecursive causal model of psychotherapy effectiveness may well be something too various and dynamic to justifiably assume to be uniform across patients or even across sessions within any given case. This means that the values initially taken by actual outcome-relevant treatment, therapist, and patient variables in a case cannot in general be counted upon to be the values that these variables strictly maintain or even average over the whole episode of treatment of that case. This is something that needs to be studied empirically, processually case by case, and settled before any causal model can properly be assumed to hold across sets of cases or over a whole episode of therapy for any single case.

THE EFFECTIVENESS OF AN EPISODE OF PSYCHOTHERAPY

A primary clinical reason for studying psychotherapy outcomes is to learn how to help therapists to be more effective with their patients. This means, first of all, trying to learn what types of patients to refer or assign to and to be persevered with by what types of therapists to optimize outcomes, because we must try to optimally use whatever therapeutic resources are presently available. Second, it means trying to learn how to make each

therapist an adequate therapist for various specific types of patients (if not all patients), because we want to improve our therapeutic resources. Fortunately, what is needed to be determined here is not the size of orthogonal fractions of outcome variance but how to adequately pair therapists with patients and match types of treatment with specific pairings of types of therapists and patients. This requires us to learn just what types of therapists can be reasonably expected to be adequately effective therapists for what types of patients and to learn eventually what types and mixes of treatments can be reasonably expected to be adequately effective for what types of patients when carried out by what types of therapists.

Differences in the variants and mixes of the various types of treatment that are actually performed by particular therapists included in an RCT comparison group in their working with the particular patients assigned to them must account at least in part for how relatively effective each type of treatment *appears in the aggregate* to be for that comparison group (see, e.g., Elkin, 1999; Orlinsky & Rønnestad, 2005, pp. 3–7; Serlin, Wampold, & Levin, 2003; Siemer & Joormann, 2003). And differences in the mixes of types of patients across individual therapists must account at least in part for how relatively effective each therapist appears to be (see Crits-Christoph, Tu, & Gallop, 2003). Likewise, whatever differences in the mixes and sequences of and degrees of expertness in the types of treatment the individual therapists actually use in each case must account at least in part for how relatively effective each of them appears to be (see, e.g., Huppert et al., 2001; Krause, Lutz, & Saunders, 2007).

However, it is ultimately the *therapist's* professional responsibility to ably and intelligently manage the process of psychotherapy no matter how treatment types or mixes drift or the patient or the therapist (as therapist) may change in the course of and as part of this process. This is a crucial consideration for the proper design and analysis of outcome studies, because it means that in fact treatments as actually provided are nested within patients (cases) and patients are nested within therapists. Although the causal influences carried by each of these three kinds of entities can affect those carried by the other two, the therapist has the responsibility for managing what happens in the psychotherapy process.

The recent research literature on therapists' aggregate contribution to psychotherapy outcomes provides us an important perspective and timely warning on this point. There is evidence that has been interpreted to mean that relatively little of the variance in patient outcomes in RCTs or in ordinary psychotherapeutic practice is on average due to therapists, about 5–10%. Whatever this amount of therapist main-effect variance is in a given RCT (and it is noted to be exceeded only by that of nonspecific treatment factors: see Elkin et al. 2006; Kim, Wampold, & Bolt, 2006; Lutz et al., 2007; Wampold, 2006; Wampold & Brown, 2005), it simply cannot correctly be interpreted as a measure of therapists' contribution to the outcome of any type of psychotherapy. In fact *all* the effects of every type of psychotherapeutic treatment are in part necessarily therapist effects, because they depend at least in part on *how* the therapist actually proceeds to do psychotherapy and to engage his or her patients in the therapy process (e.g., see Baldwin, Wampold, & Imel, 2007). In other words, the effects of every type of treatment are necessarily mediated and moderated (see, e.g., Edwards & Lambert, 2007) by the therapists who enact the treatment. It is impossible for an intelligent, competent, experienced, ethical, humane psychotherapist to do psychotherapy without somewhat inflecting what she or he does (i.e., the mix of, relative emphases on, and nuances in treatment types used, and even the continuing of rather than transferring or terminating cases) and how she or he relates with patients according to the particularities of each patient (in terms, e.g., of the patient's openness, intelligence, motivation, prior experiences in or outside of treatment, self-awareness, ability to take an adequate patient role) and of each session (according to, e.g., the patient's or therapist's intruding life crises, transient physical illnesses, acting out, and ruptures in their working alliance). This is so because psychotherapy—rather like parenting, teaching, managing, negotiating, etc.—is inherently an intimate and dynamic interrelating of persons that is somewhat constrained by the treatment type(s) the therapist invokes and by the nature and state of the patient and the therapist. Amidst all of this complexity the therapist is professionally responsible for navigating and does somewhat navigate the course of treatment, somewhat unintentionally reacting to and somewhat intentionally adjusting to the exigencies of the process what he or she does and how she or he is in this process. Thus, everything that impacts upon the outcome variables during psychotherapy sessions (and

some things outside them, such as the patient's assigned or spontaneous homework, like practicing newly taught ways of dealing with stressors between sessions) ought to be, can be, and may be somewhat influenced by the therapist.

Undoubtedly some therapists are more effective in treating some types of patients (e.g., see Huppert et al., 2001; Lutz et al., 2002), and so the question of *what type* of therapist is most effective (perhaps in part through emphasizing some treatment form or combination of forms) with *what type* of patient, as these types are identifiable *pre*treatment, is very much worth addressing. The critical issue here is how to predict who will do well enough with whom, rather than how much variance in outcomes there is between therapists *in the aggregate*. Studying cases aggregatively (e.g., in terms of averages, as ANOVA requires) is sensible for clinical purposes only when what has been aggregated is sufficiently homogeneous both in the relevant predictors or causes and in the effects of interest, so we must try to disaggregate cases into such homogeneous specific sorts (see Krause & Lutz, 2006; Lutz et al., 2006) rather than simply accept whatever heterogeneity in inputs, process, or outcomes that happens and apply ANOVA regardless.

In light of the crucial role in psychotherapy of the therapist intermediating between treatment type and patient by trying to adapt and apply the former to the latter as well as to adapt the latter to the former, it is actual therapists rather than abstract treatment types that should *initially* be compared. In actual practice, patients choose, referral sources refer to, and supervisors assign cases to particular *therapists* to take intelligent responsibility for the treatment of patients. Insofar as some input types of therapists and of patients are clearly distinguishable, these types can be aggregatively compared so as to determine which types of therapists are adequately effective in treating which types of patients. Whatever types of treatment or mixes of these treatments a therapist characteristically invokes with a given type of patient *may* prove to be part of what distinguishes the effective from the ineffective types of therapists for that type of patient, but such things as just how a therapist characteristically navigates the therapy process also need to be considered in any such type of therapists (e.g., reliably constructively responsive in treating, sometimes destructively reactive in treating, quick to refer personally unsuitable patients, quick to terminate personally unsuitable patients, and too rigidly perseverant).

DISAGGREGATING CASES RATHER THAN PARTITIONING AGGREGATE OUTCOME VARIANCE

Each individual therapist's distribution of case outcomes is the logical place to *start* the data analysis on the effectiveness of psychotherapy. (a) Pooling or aggregating cases over one sample of therapists and separately over another sample of therapists to estimate a treatment contrast or component of outcome variance (i.e., between-treatment variance divided by total variance) confounds between-therapist and within-therapist information and so biases this estimate. (b) Pooling or aggregating cases within therapists to estimate an aggregate therapist component of outcome variance (i.e., between-therapist variance divided by total variance) confounds and so biases this estimate with between-patient information. These confoundings are only clinically justifiable if they amount to not more than trivial biasing or to just random noise and so are virtually not biasing at all (as random assignment is commonly but wrongly supposed to achieve in *individual* RCTs), which are possibilities that call for corroborating evidence rather than something to simply be assumed. Such evidence is unavailable so long as *any* outcome-relevant input variable supposedly equated across comparison groups by random assignment remains unmeasured.

This means that (a) aggregating of or averaging over therapists is justifiable for RCT purposes when the sets of therapists assembled within an RCT's treatment groups are in all outcome-relevant regards nearly enough equivalent across these treatment groups, but (when it actually *is* feasible) random assignment is *not* reliably or verifiably effective at achieving the required equivalence in any individual RCT (see Krause & Howard, 2003). And, were it feasible and actually successful at achieving such equivalence, at just what multivariate distribution of therapist-variables' values (or mix of therapist types) the comparison groups *are* equivalent would remain unknown (i.e., without the necessary measurements having been taken), although this is clinically essential for us to know (see Krause & Lutz, 2006). (b) Such aggregating of or averaging over the patients assigned to each therapist in an RCT is clinically justifiable only when the sets of patients assigned to each therapist are in all outcome-relevant regards nearly enough equivalent across therapists (and, again, random assignment—were it actually achieved—cannot guarantee this or let it be verified or allow us to know at what configuration of patient-variables'

values this equivalence has occurred). Simply *assuming* such equivalence as a hypothetical justification for applying ANOVA (whether by means of experiment design, or by data analysis alone, i.e., by hierarchical multiple regression) is obviously not the same as demonstrably achieving near enough equivalence over therapists around some *known* multivariate configuration of patients' initial input values (or mix of types).

In contrast, the individual therapists' outcome distributions are potentially informative for learning who treats whom adequately insofar as we know which (pretreatment or input) patient types distinguish the adequately successfully treated patients of *each individual* therapist. These may be the same or different types across therapists, different in terms of configurations of values on the same set of patient variables or on somewhat different sets of such variables for the different therapists. To deal with such possibilities will require our uniformly validly measuring numerous patient variables pretreatment and also a finer-grained kind of data analysis than merely bivariate or even multiple or canonical correlation between input and outcome variables *unless* such correlations are virtually equal to unity. This is so because for *clinical* purposes each patient matters, and so we cannot confine ourselves to dealing with averages over variation, which is what imperfect correlations are.

Having disaggregated individual therapists' case outcomes into *sub-distributions* differentiated according to patient type, these can be compared across therapists for how effective different therapists are for each such type of patient. Some therapists may be adequately successful with a given type of patient and other therapists not, and insofar as this is the case we need to try to define what configurations of values on *therapist* variables distinguish the adequately successful from the inadequately successful types of therapists *for each already distinguished outcome-relevant type of patient*. Comparative discriminant function analyses can serve as a rough indicator of what set of therapist initial-input variables (if any) may prove to be good enough at distinguishing those therapists who are adequately successful with each given type of patient from those therapists who are not (e.g., Cohen & Cohen, 1983, pp. 458–462), but for our clinical purposes deterministic or nonstatistical methods are ultimately called for (e.g., Krause & Howard, 2002). All of this depends upon having valid measurements on the right variables for making such distinctions, which remains to be empirically settled, so there is little more to be usefully said about this now.

The comparison of the outcomes for the same homogeneous types of patients *across* individual therapists and types of therapists is most informatively done by comparing these therapists' actual *distributions* of outcomes for the same type of patient, rather than by comparing merely the means of these distributions. This is so because only when one outcome distribution does not *overlap* another is the superiority of the one over the other clinically unambiguous. Insofar as they do overlap, no matter how great their mean difference, it is possible that the therapist who *averages* less effective can nevertheless have been more effective with *some* cases of the given type (Howard, Krause, & Vessey, 1994). So it is clinically important to distinguish for what specific subtypes of patients such interaction effects may be true. This approach is essential if we hope to optimally match patients to therapists and also if we hope to identify those particular therapists whose work should be studied for how best to work with each particular type (and subtype) of patient. In this way we can best hope to define treatment types or mixes of types that are, when practiced by whom with whom, adequately effective in actual practice.[7] To presently be fixated on estimating ANOVA main effects in the design and analysis of psychotherapy RCTs is simply scientifically imprudent.

CONCLUSIONS

What is predictive need not also be causal, so certain treatment, therapist, or patient input variables' initial or pretreatment values *may* be predictive of outcomes even if these outcomes are actually causally influenced by other input or process variables rather than by or only by these predictors. Where we deal with nonrecursive causal systems, as in psychotherapy and other interpersonal situations, the control that is meant to be exercised over such systems for achieving particular results rationally must be ongoing and responsively adjustable rather than only an initial setting of inputs. But this does not prevent us from trying to take advantage of whatever predictability we can discern, and consistency in any particular therapist's effectiveness (or ineffectiveness) over some set of cases involving a particular initial type of patient provides evidence of such predictability. Such consistency may be predictable from, for example, a therapist's *already demonstrated* mastery in the performance of some type of treatment, his personality or ability characteristics, her

accumulated learning experience with clients of the given type, his or her therapy-process management style, or from some combination of qualities.

An alternative approach, that of studying average outcomes over supposedly equivalently outcome-influential sets of therapists nested in supposedly pure treatment types and with supposedly equivalently outcome-influential sets of patients nested in each of these therapists, is the traditional ANOVA approach. This ANOVA approach is vulnerable to nonrecursiveness in its partitioning of outcome variance, so any apparent effects of initial treatment types are dynamically mediated and moderated in the therapy process by some of the initial or emergent qualities of the therapists who enact these treatments and of the patients who participate in them. This makes the separately partitioned-out components of outcome variance spuriously accounted for by the factors to which they are mathematically attributed by ANOVA.

What is clinically properly sought in our RCTs is, first, predictive specificity, that is, specifically what initial types of patients are adequately treated by specifically what initial treatment types or mix of types performed by specifically what initial types of therapists. Were we to know this we could more rationally assign particular patients to particular therapists, and we could more rationally recruit and train therapists. Such training ought to use as its principal models for each type of therapist-trainee the therapists of that type working with the types of patients with whom they are reliably adequately successful. And it ought to use as its principal manuals descriptive formulations of how such therapists work with these particular types of patients.

If it turns out that some type of treatment is reliably adequately effective over a wide span of pairings of patient and therapist types, this will naturally become apparent from the results of the fine-grained research proposed here. If this is not what turns out, then we will have learned more than if we proceeded as if there simply must be treatment main effects when in fact there are not. Perhaps here lies the answer to the infamous "Dodo Bird" phenomenon of psychotherapy outcome research (see, e.g., Wampold et al., 1997).

NOTES

1. A type is meant here as something defined in terms of some set of variables: as a configuration of values, one such value from each of these variables. If you conceive of a set of variables as defining the dimensions of a multivariate space, then each point in that space defines a type and the values on these dimensions that define this point are its coordinates in this space and so also define this type.

2. Stochastic equivalence is, in one sense, equivalence as averaged out at the mathematical limit of all possible random assignments, but it does not entail equivalence for any given assignment or set of assignments before that limit is reached. It entails only a statistical tendency to probabilistically approach that limit, so any given study *may* have comparison groups quite far from equivalent on the uncontrolled and unmeasured causally relevant variables, and until we have validly measured these variables there is no way to detect how far from equivalent. See, e.g., Krause and Howard (2003).

3. A treatment manual can offer alternative ways to deal with various exigencies or can otherwise provide flexibility within the prescribed strategic approach for therapists' translating it into practice (Kendall & Beidas, 2007), but the more it does so the more conceptually ambiguous can become the treatment it defines. Furthermore, individual therapists can differ in their overall use of this flexibility and in how they do so with different types of patients. What therapist tactics a manual allows through its built-in flexibility to be inter-substitutable are not made identical in their effects on patients simply by being so allowed. This means that what are the given flexible treatment's actual effects on patients is ultimately an uncontrolled empirical matter rather than something that is definitionally-prescriptively controlled. Furthermore, to sufficiently define a type of treatment in a manual, whatever that treatment does *not* include must be explicitly precluded (see Krause, Lutz, & Saunders, 2007). So, for example, if a purely behavioral treatment is to be fairly evaluated it must preclude all therapist actions (advertent and inadvertent) that represent other types of treatments, e.g., that could have the effect of changing patients' outcome-relevant beliefs, impulse-defense dynamics, unconditional self-regard, etc. Otherwise any such actions will confound the effects of the behavioral treatment on the outcome variables insofar as these alien actions on patients do influence the outcome variables, i.e., insofar as the types of treatment represented by these actions are in fact effective.

4. The nature of the psychological situation that exists for a patient and a therapist over their sessions together, or even from moment to moment of a single session of psychotherapy, cannot reasonably be assumed to be something that is stable and uniform, nor can the psychological state (and so nature) of either the patient or the therapist from session to session. The social psychology literature on personality and situation is worth looking into in this regard (e.g., Apter & Heskin, 2001;

Ekehammar, 1974; Fleeson, 2004; Jones, 1985; Krahé, 1986; Magnusson & Ekehammar, 1978; Mischel & Shoda, 1995; Murtha, Kanfer, & Ackerman, 1996).

5. Meta-analysis as currently done does not get us around this problem. The average effect size or percentage of outcome variance in a meta-analysis for a *treatment* type—such as cognitive behavioral therapy—over a set of replicate RCTs involving that treatment type in pure form (see Krause, Lutz, & Saunders, 2007) is only a meaningful statistic insofar as the set of replicates is somehow a representative sample from some well-defined population of such replicates. Two most obviously interesting defining aspects of any such population are therapist and patient types. An average effect size for a treatment is meaningful only if the sampling error for the therapist types and patient types is indeed random error (i.e., nets out to zero), so that the average treatment effect size is not biased by therapist or patient effects (Howard et al., 1997). Has any meta-analysis of treatments' efficacy ever demonstrably achieved such representative or even random sampling? Without such sampling the individual replicate-RCTs' effect sizes reflect in part the systematic differences in the samplings of therapists and patients among these RCTs, which makes their average effect size a meaningless statistic. Similarly, for any meta-analytic estimate of an aggregate *therapist* effect, the sampling must be representative or at least random over treatment and patient types if an average therapist effect is to be meaningful. Simply pooling more studies is not tantamount to achieving the necessary sampling.

6. For a detailed critical discussion of random assignment, see Krause and Howard (2003; also see Hsu, 1989; Rosenberger & Lachin, 2002, pp. 65–73). With regard to matching, see, e.g., Anderson et al. (1980, pp. 69–112).

7. Because therapists can change over their careers in how they work and because the nature of the populations of therapists and patients can change over time, it cannot reasonably be assumed that therapists' caseload mixes or case-outcome distributions for each type of patient will remain the same, so these distributions, the typologies, and the comparisons will require recurrent updating. How and how much a therapist is changed as a therapist by sheer personal development over time or by the nature of the particular sequence of cases she or he has dealt with is an open empirical question. Insofar as a therapist is changed by the cases she or he has had, the changes are crossover effects (see, e.g., Cox, 1958, pp. 269–278) from the impact upon this therapist of some prior cases (e.g., due to a patient's suicide, a malpractice suit, a stunning breakthrough in a case) upon the impact this therapist then has upon some subsequent cases' outcomes. Such covariates complicate the patients-nested-in-therapists design by requiring the distinguishing of such phases in a therapist's way of relating to patients and of doing therapy

over any given set of temporally sequenced cases. There is a great deal more to learn about all this, especially since much more than experiences in doing psychotherapy may affect how a therapist develops as a therapist over his or her career (see, e.g., Orlinsky & Rønnestad, 2005, pp. 101–158).

REFERENCES

Anderson, S., Auquier, A., Hauck, W. W., Oakes, D., Vandaele, W., & Weisberg, H. I. (1980). *Statistical methods for comparative studies: Techniques for bias reduction.* New York: Wiley.

Apter, M. J., & Heskin, K. (2001). Basic research on reversal theory. In M. J. Apter (Ed.), *Motivational styles in everyday life: A guide to reversal theory* (pp. 77–96). Washington, DC: American Psychological Association.

Baldwin, S. A., Wampold, B. E., & Imel, Z. E. (2007). Untangling the alliance-outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology*, *75*, 842–852.

Berry, W. D. (1984). *Nonrecursive causal models.* Beverly Hills, CA: Sage.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cox, D. R. (1958). *Planning of experiments.* New York: Wiley.

Crits-Christoph, P., Tu, X., & Gallop, R. (2003). Therapist as fixed versus random effects—some statistical and conceptual issues: A comment on Siemer & Joormann (2003). *Psychological Methods*, *8*, 518–523.

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods*, *12*, 1–22.

Ekehammar, B. (1974). Interactionism in personality from a historical perspective. *Psychological Bulletin*, *81*, 1026–1048.

Elkin, I. (1999). A major dilemma in psychotherapy outcome research: Disentangling therapists from therapies. *Clinical Psychology: Science and Practice*, *6*, 10–32.

Elkin, I., Falconnier, L., Martinovich, Z., & Mahoney, C. (2006). Therapist effects in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Psychotherapy Research*, *16*, 144–160.

Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*, *13*, 83–87.

Howard, K. I., Krause, M. S., Saunders, S. M., & Kopta, S. M. (1997). Trials and tribulations in the meta-analysis of treatment differences: Comment on Wampold et al. (1997). *Psychological Bulletin*, *122*, 221–225.

Howard, K. I., Krause, M. S., & Vessey, J. (1994). Analyzing clinical trial data: The problem of outcome overlap. *Psychotherapy*, *31*, 302–307.

Hsu, L. M. (1989). Random sampling, randomization, and equivalence of contrasted groups in psychotherapy outcome research. *Journal of Consulting and Clinical Psychology*, *57*, 131–137.

Huppert, J. D., Butka, L. F., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2001). Therapists, therapist variables, and cognitive behavioral therapy outcomes in a multicenter trial for panic disorder. *Journal of Consulting and Clinical Psychology*, *69*, 747–755.

Jones, E. E. (1985). Major developments in social psychology during the past five decades. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (3rd ed., pp. 47–107). New York: Random House.

Kendall, P. C., & Beidas, R. S. (2007). Smoothing the trail for dissemination of evidence-based practices for youth: Flexibility within fidelity. *Professional Psychology: Research and Practice*, *38*, 13–20.

Kim, D., Wampold, B. E., & Bolt, D. M. (2006). Therapist effects in psychotherapy: A random effects modeling of the NIMH TDCRP data. *Psychotherapy Research*, *16*, 161–172.

Krahé, B. (1986). Similar perceptions, similar reactions: An idiographic approach to cross-situational coherence. *Journal of Research in Personality*, *20*, 349–361.

Krause, M. S., & Howard, K. I. (2002). The linear model is a very special case: How to explore data for their full clinical implications. *Psychotherapy Research*, *12*, 475–490.

Krause, M. S., & Howard, K. I. (2003). What random assignment does and does not do. *Journal of Clinical Psychology*, *59*, 751–766.

Krause, M. S., & Lutz, W. (2006). How we really ought to be comparing treatments for clinical purposes. *Psychotherapy*, *43*, 359–361.

Krause, M. S., Lutz, W., & Saunders, S. M. (2007). Empirically certified treatments or therapists: The issue of separability. *Psychotherapy*, *44*, 347–353.

Lutz, W., Leon, S. C., Martinovich, Z., Lyons, J. S., & Stiles, W. B. (2007). Therapist effects in outpatient therapy: A three-level growth curve approach. *Journal of Counseling Psychology*, *54*, 32–39.

Lutz, W., Martinovich, Z., Howard, K. I., & Leon, S. C. (2002). Outcomes management, expected treatment response, and severity-adjusted provider profiling in outpatient psychotherapy. *Journal of Clinical Psychology*, *58*, 1291–1304.

Lutz, W., Saunders, S. M., Leon, S. C., Martinovich, Z., Kosfelder, J., Schulte, D., et al. (2006). Empirical and clinical useful decision making in psychotherapy: Differential predictions with treatment response models. *Psychological Assessment*, *18*(2), 133–141.

Magnusson, D., & Ekehammar, B. (1978). Similar situations—similar behaviors? A study of the intraindividual congruence between situation perception and situation reactions. *Journal of Research in Personality*, *12*, 41–48.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*, 246–268.

Murtha, T. C., Kanfer, R., & Ackerman, P. L. (1996). Toward an interactionist taxonomy of personality and situations: An integrative situational-dispositional representation of personality traits. *Journal of Personality and Social Psychology*, *71*, 193–207.

Orlinsky, D. E., & Rønnestad, M. H. (2005). *How psychotherapists develop: A study of therapeutic work and professional growth*. Washington, DC: American Psychological Association.

Rosenberger, W. F., & Lachin, J. M. (2002). *Randomization in clinical trials: Theory and practice*. New York: Wiley.

Serlin, R. C., Wampold, B. E., & Levin, J. R. (2003). Should providers of treatment be regarded as a random factor? If it ain't broke, don't "fix" it: A comment on Siemer & Joormann (2003). *Psychological Methods*, *8*, 524–534.

Siemer, M., & Joormann, J. (2003). Power and measures of effect size in analysis of variance with fixed versus random nested factors. *Psychological Methods*, *8*, 497–517.

Wampold, B. E. (2006). The psychotherapist. In J. C. Norcross, L. E. Beutler, & R. F. Levant (Eds.), *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions* (pp. 200–208, 236–238). Washington, DC: APA.

Wampold, B. E., & Brown, G. S. (2005). Therapist variability in practice. *Journal of Consulting and Clinical Psychology*, *73*, 914–923.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes." *Psychological Bulletin*, *122*, 203–215.