Understanding processes of change: How some patients reveal more than others – and some groups of therapists less – about what matters in psychotherapy

Robert J. DeRubeis[a], **Lois A. Gelfand**[a], **Ramaris E. German**[a], **Jay C. Fournier**[b], and **Nicholas R. Forand**[c]

[a]Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

[b]Department of Psychiatry, University of Pittsburgh School of Medicine, PA, USA

[c]Department of Psychiatry, The Ohio State University Wexner Medical Center, OH, USA

Abstract

We identify difficulties researchers encounter in psychotherapy process-outcome investigations, and we describe several limitations of the popular "variance accounted for" approach to understanding the effects of psychotherapy. Using data simulations, we show how the expected correlation between an excellent measure of therapy quality and outcome would be surprisingly small (approximately .25) under conditions likely to be common in psychotherapy research. Even when we modeled conditions designed to increase the likelihood that strong process-outcome relationships would be observed, we found that the expected correlations were still only in the modest range (.38 – .51). We discuss the implications of our analysis for the interpretation of process-outcome findings as well as for design considerations in future investigations.

Keywords

Psychotherapy; Process-outcome Research; Mechanisms of Change; Common factors; Technique

---

Although most readers of the psychotherapy research literature will agree with the broad claim that "psychotherapy works," there is little agreement concerning the relative importance of the elements that constitute effective psychotherapy. Some of the obstacles to resolving these disagreements are well known; these include the difficulty of defining and operationalizing the elements of therapy, as well as measuring or manipulating them. However, we believe there are additional issues that have thus far received little attention, but nevertheless might have a substantial impact on estimates of the relationships between processes and outcomes.

---

[1]Skew = −.45; where negative values indicate left-skewed distributions and positive values indicate right-skewed distributions. A value of zero indicates a symmetric distribution.

[2]Kurtosis = 1.7 (heavy tailed), where a normal distribution has a kurtosis of 0. Positive values indicate heavy tailed distributions and negative values indicate light-tailed distribution.

The "variance accounted for" approach to characterizing the sources of change in psychotherapy

We begin by offering a critical assessment of the dominant method used to illustrate these relationships. Lambert and colleagues have published a series of influential papers in which they attempt to characterize in quantitative terms the effects of various psychotherapy factors on patient outcomes. Their approach is to divide overall variance in psychotherapy outcomes into the percentages attributable to various therapeutic factors (Asay & Lambert, 1999; Lambert, 1992, 2013; Lambert & Barley, 2001; Lambert & Barley, 2002). In their framework, typically represented in a pie chart, "Common Factors" and (specific) "Techniques" constitute the factors that vary across psychotherapy cases and are presumed to be under the control of the therapist. Also included in the pie chart are two other factors, "Expectancy" and "Extratherapeutic Change" (or "Client/Life"). Collectively, these four factors are said to account for the total variation in outcome that results from psychotherapy. The authors attribute 30% of outcome variance to common factors. (The therapeutic alliance is considered to be one of the common factors.) Techniques are said to account for 15%, and the remaining variance is attributed to extratherapeutic change (40%) and expectancy (15%).

The simplicity and seeming comprehensiveness of this way of understanding change has evidently held great appeal among those who publish on process and outcome research in psychotherapy; according to a "Google scholar" search, the publications in which the pie charts appear have been cited over 1,500 times. Commonly, those citing the papers refer to the pie charts as if they represented quantitative reviews (meta-analyses) of the psychotherapy outcome literatures. However, as Lambert and Barley (2002) state, the charts are "not derived directly from meta-analytic techniques, (but instead) characterize the research findings of a wide range of treatments, disorders, and ways of measuring client and therapist characteristics (p. 18)." There is, therefore, no obvious way to replicate or verify the conclusions summarized in the pie charts. Moreover, as Lambert (2013) states, "It is really difficult to partition and compare sources of variability in psychotherapy because no single study encapsulates all the variables of interest" (p. 200).

Another, perhaps less well understood limitation of Lambert's (1992) approach concerns the use of the "percent variance accounted for" concept to indicate the relative contribution to psychotherapeutic change of a given factor. This metric is informative only insofar as there is substantial variability in each of the factors in each of the constituent studies. To illustrate, imagine that in a given population of therapist-patient dyads, a high level of technical skill is required if positive outcomes are to result from the therapy. If in a study of dyads from that population all therapists possess a high level of technical skill, the variance accounted for by technical skill will be very small, even though technical skill is critical for treatment success. Therefore, any factor on which the therapists did vary could appear to be important and, crucially, more important than skill. Although Lambert and Barley (2002) point to the fact that the studies they summarized "span extremes in research designs, and are especially representative of studies that allow the greatest divergence in the variables that determine outcome (p. 18)," this is a statement about the heterogeneity of the studies, and not about the variability of the factors represented within the individual studies. Thus, when they

"average(d) the size of the contribution each predictor made to final outcome (p.18)," they were doing so across studies that, collectively, may have represented a wide range on all of the factors, even if many or most of these studies, individually, represented rather narrow ranges. If the ranges were in fact limited within the studies, averaging the findings across the studies, no matter how different they are from one another, would do nothing to address this problem.

An alternative approach to understanding the effects of therapeutic factors on outcomes

With these and other issues in mind, we developed an alternative conceptualization that we hope will lead to refinements in thinking and research concerning the elements of psychotherapeutic processes. In our framework we explicitly incorporate two phenomena that are inadequately addressed in the "variance explained" approach. The first involves a hidden moderator variable in the outcome and process-outcome literatures: variability among patients in the degree to which the quality of therapy provided to them will affect their outcomes. The second addresses the likelihood that few, if any, studies of psychotherapeutic processes include sufficient variation in potentially relevant factors (e.g., adherence to the treatment; engagement in relationship-enhancing behaviors) to allow for meaningful examinations of the causal links between therapeutic factors and outcomes. In addition, we will briefly touch upon a few other phenomena that may be useful to consider when designing psychotherapy studies and interpreting results. In the process, we hope to stimulate ideas for future endeavors that can provide more informative tests of hypotheses embedded in the question: "How does psychotherapy work?"

As psychotherapy researchers, we are interested in the relationships between various therapeutic processes and outcome. In our framework, we use the summary term *quality* to represent all therapeutic influences on improvement in therapy. It is therefore an idealized construct in that it comprises all and only those elements, variability in which accounts maximally for variability in outcomes. It includes the therapist's use of interpersonal skills to engage the patient, to avoid or address unhelpful tensions in the service of promoting a mutual commitment to therapeutic growth. In most discussions of therapy processes these elements are subsumed in the "relationship" rubric. Also included under *quality* would be the skill with which the therapist assesses the patient's needs and then determines a course of action to address those needs. The extent to which a therapist succeeds in these aims might be referred to as the "fit" of the therapist's plan with what would most help the patient. Finally, *quality* also includes the skill with which the therapist employs appropriate "techniques." This would incorporate the intensity and frequency of the therapeutic procedures (similar to the concept of adherence) as well as the competence with which they are provided. It is beyond the scope of this paper to address the ways in which relationship, fit, and technique factors might interact with each other, or how they might each contribute in unique ways to therapeutic change. The conclusions we derive using the *quality* construct could, in future research, be applied to these or other factors.

A second causal influence, which we will assume is independent of therapy *quality*, concerns the variability in the potential for patients to improve with the passage of time,

irrespective of the quality of therapy provided to them. Hereafter we will refer to the influence of the passage of time on a given patient's outcome as that patient's *time-effect*. The *time-effect* construct is similar to Lambert's Expectancy factor.

In the following, *quality* and the *time-effect* are the only influences on therapy outcome that we model. We will not include random extratherapeutic events as, by definition, these random events are not attributable to the therapist, his or her therapeutic methods, or the patient; these would present as random error or unexplained variance in any model. We note that insofar as positive (or negative) life events result directly from the therapeutic process and go on to affect outcome (such as when problem-solving results in changes in the patient's environment), they would be accounted for by the therapist-generated factors and thus captured by the *quality* variable in statistical models.

Taking into account Variability in Patients and Therapists

Correlational analyses assess the degree of linear relation between two variables. The theoretical maximum positive correlation is 1.0, but the maximum correlation attainable in any study is usually less than 1.0. It is well understood that the unreliability of measures, as well as the occurrence of events outside the explanatory system (such as random events) constrain the magnitude of the relationships that will be observed in process-outcome research. However, two other phenomena that can limit the magnitude of observed relations between process and outcome variables have tended to be overlooked; these are the focus of our analysis. One of these concerns the variability in *quality* within individual process-outcome studies. The other concerns variability across patients in how, and to what degree, they would respond to variation in therapy *quality*. In the following we illustrate how large a role these phenomena can play in the determination of estimates of process-outcome relationships in typical psychotherapy studies.

Figure 1 depicts a variety of relationships that might exist between percent improvement (hereafter *improvement*) and *quality*, where *improvement* can range from 0% (none) to 100% (complete resolution of relevant problems). For simplicity, we do not consider negative values, which would represent deterioration. *Quality* is scaled to vary from 0%, reflecting the absence of any therapeutic elements (as in, for example, a wait-list condition), to 100%, in which therapeutic factors are optimal. For ease of discussion, we refer to the 0%, 25%, 50%, 75%, and 100% points as Absent, Modest, Good, Excellent, and Highest *quality*. A linear relationship with a slope of 1.0 is what would be observed, assuming perfectly valid and reliable measurement of both *quality* and *improvement*, if all patients would improve as a direct one-to-one function of the level of therapy *quality* provided to them. We will refer to a patient or set of patients who exhibit this "Response Pattern" as "Pliant," in that their response to therapy will be excellent if excellent therapy is provided, poor if the therapy is poor, etc. These concepts are instantiated in the framework we employ throughout the paper, such that when *quality* is Highest, the Pliant patient will improve to the fullest extent possible (i.e., 100% *improvement*), and gradations of *quality* between Absent and Highest engender *improvement* proportional to the degree of *quality*. For Pliant patients, then, assuming perfect measurement, the correlation between *quality* and *improvement* would be 1.0.

While it is theoretically possible that all patients are Pliant, it is more in line with our clinical impressions to represent patients as differing in their patterns of response to variations in *quality*. Figure 1 includes depictions of the relationship between *quality* and *improvement* for eight types of patients who are not Pliant. We name four of these Response Patterns: the "Spontaneous Remitter," who experiences maximum improvement irrespective of *quality*; the "Easy" patient, who would experience considerable improvement even without therapy, and would evidence maximal improvement if *quality* is at least Good; the "Challenging" patient, who would experience no improvement unless *quality* is at least Good, and would achieve less than maximal improvement even if optimal therapy were provided; and the "Intractable" patient, who would experience no improvement no matter how low or high the *quality* of therapy provided to him or her.

The patients in this conceptualization are represented on a continuum such that the relationship between *quality* and *improvement* is always nonlinear (with the exception of the Pliant patients, Intractable patients, and Spontaneous Remitters) in the following way. Each relationship has a "hockey stick" shape composed of a horizontal piece, such that there is no relationship between *quality* and *improvement* over that range of *quality*, and a diagonal piece, where there is a linear relationship between *quality* and *improvement* (slope = 1.0) through that part of the range. Response Patterns are created by applying floor and ceiling effects to the diagonal lines such that a Response Pattern is uniquely identified by the Y-intercept of the line before any floor or ceiling effect is applied. These intercepts range from −100% ("Intractable") to +100% ("Spontaneous Remitter"). (Note: Figure 1 depicts the 5 named Response Patterns, plus 4 intermediate Response Patterns that are not associated with specific labels.) The horizontal piece may occur at only at either 100% improvement (a ceiling effect, as is true for the Easy patient) or 0% improvement (a floor effect, as for the Challenging patient). The juncture between the horizontal and diagonal piece may occur at any value of *quality*. An example, from Figure 1, is the Easy patient, whose diagonal piece begins at 50% improvement, showing that even with the lowest quality of therapy (therapy *quality* is Absent), he or she will improve substantially, but not maximally. For the Easy patient, the diagonal piece reaches 100% when *quality* is Good. The horizontal piece, which runs from Good to Highest, represents the idea that at any level of *quality* Good or higher, the Easy patient will experience full (100%) improvement.

Figure 2 (below) presents a different view of the model introduced in Figure 1, focusing on five of the patient Response Patterns and showing how the *time-effect* and *quality* influence *improvement* in different patients. When *quality* is Absent (0%), any *improvement* is attributed only to a patient's *time-effect*. As *quality* increases, the magnitude of the *time-effect* remains constant; at higher levels of *quality*, all and only the *improvement* beyond that attributable to the *time-effect* is due to *quality*. Note especially how *quality* can play either a crucial role (Pliant Patient), no role at all (Intractable Patient or Spontaneous Remitter), or degrees in between (Challenging Patient and Easy Patient).

This model is only one example of how patients might vary in their response to therapy quality, but two of its features would likely be represented in any realistic depiction of the relationships between *quality* and *improvement*. First, for some patients (e.g., Spontaneous Remitters and Intractable patients), there is little or no relationship between *quality* and

improvement. Second, the relationship between *quality* and *improvement* is not linear for most patients. These two features will serve to constrain estimates of the linear relationship between *quality* and *improvement* in any process-outcome study, because no process-outcome study would ever contain only Pliant patients.

We performed a set of simulations in order to estimate the correlation between *quality* and *improvement* under a variety of conditions. In each simulation, we took 1,000 random draws from a distribution of the Response Patterns and crossed them with random draws from a distribution of *quality* values. Each pairing of a level of *quality* with a Response Pattern resulted in a percentage value reflecting the level of *improvement* that would be expected for that particular therapy case. We then calculated estimates of the correlation between *quality* and *improvement*, using all 1,000 cases in the simulation. The value of the dependent variable, *improvement*, was determined by *quality* and Response Pattern, as described above and exemplified in the following three cases (see Figure 1 or Figure 2). If Response Pattern is Spontaneous Remitter and *quality* is Excellent (75), *improvement* is 100%. If Response Pattern is Easy and *quality* is Modest (25), *improvement* is 75%. If Response Pattern is Pliant and *quality* is Good (50), *improvement* is 50%. In this set of three cases, the correlation between *quality* and *improvement* is .50.

In the base simulation, we drew *quality* levels and Response Patterns from uniform distributions, representing a situation in which each of the levels of *quality*, and each of the Response Patterns, was equally likely. In this simulation, the correlation between *quality* and *improvement*, assuming perfect measurement, was .33 (Case 1, Table 1). Recall that *quality* is an idealized construct. As such, its validity coefficient is 1.0.

Distribution of Therapy Quality and Patient Response Patterns in Studies

In many process studies it is the intention of the investigator and the therapists to provide high-quality treatment. Thus, we assume it is unrealistic to represent variation in *quality* with a uniform distribution. In our next simulation, we therefore drew *quality* from a left-skewed distribution[11] , depicted in Figure 3, in which the bins that contain Good and Excellent *quality* were over-represented, compared to those that contain Absent, Modest, and Highest. In this simulation we retained a uniform distribution for Response Pattern. The resulting correlation between *quality* and *improvement*, assuming perfect measurement, was .18 (Case 2, Table 1).

In each of the two simulations we have described thus far we have represented the different Response Patterns equally across the range. Since it is possible or even likely that a uniform distribution does not represent samples or populations of patients who participate in process studies, we conducted a simulation that reflected a more realistic distribution of Response Patterns. The distribution we chose was like that of the left-skewed *quality* distribution, such that the extreme patterns (Spontaneous Remitter and Intractable) were represented less frequently than were the less extreme patterns. Likewise, Response Patterns between Intractable and Challenging were relatively rare. The correlation we obtained with this simulation, .39 (Case 3, Table 1), represents an optimistic estimate of the correlation a researcher would obtain, with perfect measurement, under what is arguably the most realistic pair of distributions among those we tested.

As can be seen from Table 1, for a given distribution of *quality* the correlation between *quality* and *improvement* increases as the proportion of Pliant or near-Pliant cases increases. Note that in an idealized case in which the sample is composed entirely of Pliant patients and the validity coefficient of *quality* and *improvement* is 1.0, the correlation will be 1.0. We considered a case in which the Response Patterns represented are predominantly in the middle of the range, centered around Pliant (hereafter referred to as Predominantly Pliant; see Figure 4[22]) in order to demonstrate how correlation estimates can be increased by focusing on such patients in process-outcome studies. The Predominantly Pliant distribution was represented in a simulation that used a skewed distribution of *quality*; the resulting correlation was .59 (Case 4, Table 1), assuming perfect measurement. When we substituted a uniform *quality* distribution for the skewed distribution, we obtained the highest correlation of any of the five simulations, .79 (Case 5, Table 1). This shows how important it is to understand and – to the degree possible and ethically justifiable – influence the distributions of both *quality* and Response Patterns in studies relating therapy processes to outcomes. Across the simulations, correlations ranged from .18 (Case 2), when the *quality* distribution was skewed and the Response Pattern distribution was uniform, to .79 (Case 5), when *quality* was uniform and Response Patterns were Predominantly Pliant. Case 5 reflects distributions that would be advantageous to obtain in order to reveal causal relations between therapy processes and outcomes.

In the simulations described thus far, perfect measurement validity was assumed. The correlations from these simulations are represented in the first column of numbers in Table 1. The second column of numbers contains the estimates for each of these correlations under the assumption that the validity coefficients for both the independent variable (*quality*) and the dependent variable (*improvement*) are .80. We provide these adjusted correlations as a reminder that, in addition to the problems we have already addressed, the reliability and validity of the measures can place further constraints on the magnitude of the correlations one can obtain in empirical studies of these phenomena. A validity coefficient of .80 for *improvement* would seem to be achievable, but our guess is that .80 is higher than the validity of even the best psychotherapy process measures, if for no other reason than that the validity coefficient is limited by the reliability of the measure. The resulting correlations (aside from the idealized case) range from .12 to .51. These coefficients, then, represent what are likely to be the upper bounds on findings that relate processes to outcomes. Insofar as a process measure does not capture all of the elements of the therapy process that contribute to change, the expected correlations would be even lower.

The models we tested do not take into account the effects of factors or events that are outside of the therapist's control and not related to Response Patterns, such as an unexpected illness in the family, or a contact initiated by a long lost friend. Lambert (1992) includes these events in his term, "extratherapeutic change," to which he attributes 40% of outcome variance. Although these uncontrollable factors may not be as influential, on average, as the 40% figure would suggest, their effects on *improvement* cannot be assumed to be negligible. To the extent that these events do affect *improvement*, each of the correlations in Table 1 is an overestimate of the relationship that would be obtained between *quality* and *improvement*.

Implications and Considerations for Future Investigations

In the preceding we showed how, in a process-outcome study, the finding of a relation between a causal process and change can depend as much upon the distribution of *quality* and on the make-up of a patient sample as it does on the capacity of the process to promote change. Findings regarding individual therapeutic factors will be similarly affected. For example, if in one study, therapists vary considerably in their ability to establish a good therapeutic relationship, a large and significant effect might be obtained between the measure of the relationship and outcome. If in a replication, all therapists establish good to excellent relationships with their patients, an association between relationship and outcome would likely not be observed. Investigations that aim to characterize the relative importance of two therapy processes, such as therapist adherence to treatment and the therapeutic relationship, can be subject to similar confounds. Even if two processes could be considered equally important in some contexts, a given study in which they are compared could yield results that seem to show that one of them is far more important than the other. For example, the relationship factor might appear to have a stronger association with outcome, relative to the adherence-outcome association, if the therapists vary substantially in their ability to foster a positive relationship but are all similar to each other in their adherence to the treatment.

Attending to the nature of the therapist sample

We have shown that in therapy process studies, the strength of process-outcome associations will be constrained insofar as the range of therapy quality is restricted. One implication of this is that information available about the therapist sample should be considered when evaluating the usefulness of the study in accessing processed outcome associations. Another implication is that investigators could consider these features of the sample of therapists and patients who participate in their studies. To the extent it is feasible and ethical, samples of therapists could be identified that represent a broad range on the therapy variables of interest.

Alternatively, process variables can be manipulated experimentally, such as in training studies, in which therapists are assigned randomly to participate, or not participate, in training that emphasizes a potentially important therapy process (Crits-Christoph, Connolly Gibbons, Crits-Christoph, Narducci, Schamberger, & Gallop, 2006; Wiles et al., 2013). In an ideal study comparing two processes, therapists would be randomly assigned to a condition without training, a condition with training emphasizing one of the two processes, a condition involving training emphasizing the other process, and a fourth condition in which the two processes are equally emphasized.

Attending to the nature of the patient sample

A finding that a therapeutic process accounts for change is, as we have shown, also partially dependent on the nature of the patient sample. Insofar as a sample is composed of patients whose Response Patterns diverge from Pliant, relationships between a process and *improvement* will be obscured. This same principle holds for treatment comparisons.

Consider a study with two active treatments and a control condition, where one of the treatments (the "strong" treatment) is able to address the problems presented by more difficult cases, as well as less difficult ones. The "weak" treatment is as effective with less difficult cases, but not as effective with difficult ones. The weak treatment would correspond in our system to a Good *quality* level, whereas the strong treatment would correspond to an Excellent or Highest level of *quality*. Consulting the relationship between *quality* and *improvement* for the Easy patients represented in Figure 2 is illustrative. In a sample composed entirely of Easy patients, the Good treatment will result in 100% *improvement*, and it will lead to much greater *improvement* than will a wait-list control condition (where *quality* is Absent or Modest). An Excellent/Highest *quality* treatment will fare no better than a Good treatment in this sample of patients. This pattern of findings would support the inference that neither treatment is more effective than the other, and that both treatments outperform a control condition. If the study were conducted instead with a sample of Challenging patients, a very different pattern of findings would be obtained. The Excellent/Highest quality treatment would prove superior to the Good treatment, and the Good treatment would not outperform the control condition.

A recent meta-analysis provides an example of how a treatment can be shown to be superior, or not superior, to a control condition, depending on the nature of the sample. The beneficial effects of psychotherapy, relative to placebo, were evident in samples of patients with "high" levels of depressive symptoms, but not in samples of patients with "low" levels of severity (Driessen, Cuijpers, Hollon, & Dekker, 2010). Although the categories "low severity" and "Easy" are not the same, it seems likely that individuals with milder depressions may benefit from a less potent intervention and will not benefit substantially more from a more potent one, which is how we have represented the "Easy" patient.

"Golden zones"

The implications are that the nature of the patient sample should be considered when interpreting research or when designing studies. Whereas therapist behaviors can be measured and, to some degree, experimentally manipulated, Response Patterns cannot be measured directly, nor can they be manipulated. This makes it difficult to control the distribution of Response Patterns represented in a sample of patients. Our intuitions tell us that patients vary considerably in the extent to which they require – as well as in how much they would benefit from – high quality psychotherapy. Attempting to make these judgments after therapy has concluded would confound outcome with predictors of outcome. Although it seems unlikely that a researcher could ever identify with great precision Response Patterns *a priori*, some patient characteristics indicate a high likelihood that a patient's symptoms will remit spontaneously or that they will not respond to therapy. Characteristics of likely spontaneous remitters include: a) being in a first episode of depression; b) a brief episode duration; and c) mild to moderate symptom levels. Characteristics at the other extreme on some of these same dimensions (i.e., chronic course, severe symptoms, and "treatment resistant") likely indicate patients who will be *intractable* or *challenging*. Excluding patients who appear to be at the extremes in regard to Response Patterns would result in a sample that is similar to the one we modeled as Predominantly Pliant, and therefore would include a

relatively low percentage of patients for whom important therapy processes have little to do with their degree of improvement during therapy.

The pharmaceutical industry has taken this approach in the design of placebo-controlled trials. Patients who are expected to be placebo responders are excluded, as are those whose symptoms are so severe that they are not expected to exhibit a good response to active medication. The range in between these two groups, which is the focus of pharmaceutical trials, has been termed the "golden zone" (Thase, 1999; see also Fournier et al., 2010). In our scheme, an analogous effort would result in a focus, to the extent possible and ethical, on the Pliant, and to a lesser extent, the Easy and Challenging patients, to the exclusion of Spontaneous Remitters and Intractable patients. As discussed earlier, there are indications in the outcome literature on psychotherapy for depression that severity might also serve well as a patient classifying variable in efforts to distinguish active from control psychotherapies (Driessen et al., 2010), suggesting that severity should be considered as a dimension on which to select patients for process-outcome studies.

Using meaningful units in process and outcome measures

As we have shown, both the means and the variabilities of the process measures in a study sample need to be accounted for when drawing inferences from their observed associations with outcome. Likewise, comparisons of findings between investigations should take into account any differences in the relevant distributions. The approach we have presented would work best to the extent that measures of *quality* components and *improvement* could be scaled with meaningful units. Arguably, there are outcome measures currently employed for which this is true. For example, the Beck Depression Inventory (BDI-II; Beck, Steer, & Brown, 1996) and the Hamilton Rating Scale for Depression (Hamilton, 1960) are measures for which raw scores are interpretable, with meaningful 0-scores, cutoff scores for mild, moderate, and severe symptom levels, and consensus on the interval that defines a "clinically meaningful" change. We employed percent improvement as our outcome metric in this paper, but the ideas we presented could be adapted to other change metrics, such as a raw change score.

Measures of psychotherapeutic processes tend not to have meaningful units, but rather conform to the idea that "more is better." We recommend that attempts be made to define meaningful intervals, or anchor points, for measures of therapeutic processes. We also recommend measuring multiple process variables with commonly-used measures in any given process-outcome study. This would allow for comparisons across studies of the distributions of these measures, and it would provide the opportunity for estimates to be made of the inter-correlations of the components.

Accounting for and avoiding confounds in process-outcome investigations

We have shown that typical studies of psychotherapy can be expected to yield process-outcome correlations that are below .30, even if a powerful process has been identified and measured. In this context, with meta-analytic estimates of the alliance-outcome correlation falling in the range of .20 to .30 (Horvath, Del Re, Flückiger, & Symonds, 2011), one might be tempted to conclude that the alliance drives outcome, at least in the short term. However,

few of the alliance studies have employed designs that can disentangle cause from consequence (Barber, Khalsa, & Sharpless, 2010; DeRubeis, Brotman, & Gibbons, 2005). We are aware of eight published studies, including four from our lab, in which the investigators controlled for the critical temporal confound. When we combined the results from these studies, and weighted them by study sample size, the resulting estimate of the correlation was .11 (DeRubeis, 2012, October). This suggests that the higher estimates, such as those reported in the Horvath et al. meta-analysis, may have been obtained at least in part because better outcomes lead to better alliances.

Process-outcome designs often contain other confounds that can produce spurious findings. One example occurs when the therapists in the study are represented in multiple dyads, yet these dyads are treated as independent observations. Appropriate analyses of these kinds of data would take into account the non-independence of dyads that share the same therapist. Failing to do so ignores the fact that patients' characteristics or behaviors can influence therapist behaviors, and it may be these same patient factors that promote positive therapeutic change. Estimates of alliance-outcome (Horvath, Del Re, Flückiger, & Symonds, 2011) and adherence-outcome (e.g., DeRubeis & Feeley, 1990; Feeley, DeRubeis, & Gelfand, 1999) relationships may be inflated, therefore, due to the fact that most studies of these processes have not attended to this potential confound (but see Del Re, Flückiger, Horvath, Symonds, & Wampold, 2012, for an exception). For a more extensive discussion of these and other common confounds in process-outcome research, see Crits-Christoph, Connolly Gibbons, and Mukherjee (2013).

Focusing investigations on key moments

We have focused thus far on research methods that could reveal important processes that tend to recur across a course of psychotherapy. However, effective therapy might best be understood as comprising relatively few key moments, or critical sessions, that largely account for change in psychotherapy. This possibility was delineated by Rice and Greenberg (1984), who recommended selecting as study points those moments within a session in which a notable change is observed in one or more key dimensions of the patient's psychological functioning (see also Aspland, Llewelyn, Hardy, Barkham, & Stiles, 2008; Greenberg, 2007; Pascual-Leone, Greenberg, & Pascual-Leone, 2009). Tang and DeRubeis's (1999) "sudden gains" methodology, in which correlates of sizeable single session symptom reductions are examined, is another promising example of this type of approach (see also Tang, DeRubeis, Beberman, & Pham, 2005; Tang, DeRubeis, Hollon, Amsterdam, & Shelton, 2007).

In conclusion, we have presented a framework that, unlike the "variance accounted for" approach, takes into account the potential effects of variation in patient Response Patterns and therapy *quality*. Findings from process-outcome and comparative outcome investigations will be misinterpreted insofar as these effects are not considered. Moreover, studies designed to accommodate these effects explicitly could reveal a more nuanced understanding of the importance of psychotherapeutic processes in the facilitation of positive patient outcomes.

References

Asay, TP.; Lambert, MJ. The empirical case for the common factors in therapy: Quantitative findings. In: Hubble, MA.; Duncan, BL.; Miller, SD., editors. The heart and soul of change: What works in therapy. 1st ed.. American Psychological Association; Washington, DC: 1999. p. 23-55.

Aspland H, Llewelyn S, Hardy GE, Barkham M, Stiles W. Alliance ruptures and rupture resolution in cognitive-behavior therapy: A preliminary task analysis. Psychotherapy Research. 2008; 18:699–710. [PubMed: 18815951]

Beck, AT.; Steer, RA.; Brown, GK. Manual for the Beck Depression Inventory. 2nd ed.. The Psychological Corporation; San Antonio, TX: 1996.

Barber, JP.; Khalsa, S.; Sharpless, BA. The validity of the alliance as a predictor of psychotherapy outcome. In: Barber, JP.; Khalsa, SR.; Sharpless, BA., editors. The therapeutic alliance: An evidence-based guide to practice. Guilford Press; New York: 2010. p. 29-43.

Crits-Christoph P, Connolly Gibbons MB, Crits-Christoph K, Narducci J, Schamberger M, Gallop R. Can therapists be trained to improve their alliances? A preliminary study of alliance-fostering psychotherapy. Psychotherapy Research. 2006; 16:268–281.

Crits-Christoph, P.; Connolly Gibbons, MB.; Mukherjee, D. Psychotherapy process-outcome research. In: Lambert, MJ., editor. Bergin and Garfield's handbook of psychotherapy and behavior change. 6th ed.. John Wiley & Sons; Oxford, England: 2013. p. 298-340.

Del Re AC, Flückiger C, Horvath AO, Symonds D, Wampold BE. Therapist effects in the therapeutic alliance-outcome relationship: A restricted-maximum likelihood meta-analysis. Clinical Psychology Review. 2012; 32:642–649. [PubMed: 22922705]

DeRubeis, RJ. Mechanism of change in the treatment of depression; Incorporating Use-Inspired Basic Research into Research on Intervention Efficacy. Symposium conducted at a meeting sponsored by the National Institutes of Health (NIH) Common Fund program on the Science of Behavior Change (SOBC): Revisiting Pasteur's Quadrant: Use-Inspired Basic Research; Washington, DC. 2012, October;

DeRubeis RJ, Brotman MA, Gibbons CJ. A conceptual and methodological analysis of the nonspecifics argument. Clinical Psychology: Science and Practice. 2005; 12:174–183.

DeRubeis RJ, Feeley M. Determinants of change in cognitive therapy for depression. Cognitive Therapy and Research. 1990; 14:469–482.

Driessen E, Cuijpers P, Hollon SD, Dekker JJ. Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. Journal of Consulting and Clinical Psychology. 2010; 78:668–680. [PubMed: 20873902]

Feeley M, DeRubeis RJ, Gelfand LA. The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. Journal of Consulting and Clinical Psychology. 1999; 67:578–582. [PubMed: 10450629]

Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, Fawcett J. Antidepressant drug effects and depression severity: a patient-level meta-analysis. The Journal of the American Medical Association. 2010; 303:47–53.

Greenberg LS. A guide to conducting a task analysis of psychotherapeutic change. Psychotherapy Research. 2007; 17:15–30.

Hamilton M. A rating scale for depression. Journal of Neurology, Neurosurgery and Psychiatry. 1960; 23:56–62.

Horvath AO, Del Re AC, Flückiger C, Symonds D. Alliance in individual psychotherapy. Psychotherapy. 2011; 48:9–16. [PubMed: 21401269]

Lambert, MJ. Psychotherapy outcome research: Implications for integrative and eclectic therapists. In: Norcross, JC.; Goldfried, MR., editors. Handbook of Psychotherapy Integration. Basic Books; New York: 1992. p. 94-129.

Lambert, MJ. The efficacy and effectiveness of psychotherapy. In: Lambert, MJ., editor. Bergin and Garfield's handbook of psychotherapy and behavior change. 6th ed.. John Wiley & Sons; Oxford, England: 2013. p. 169-218.

Lambert MJ, Barley DE. Research summary on the therapeutic relationship and psychotherapy outcome. Psychotherapy. 2001; 38:357–361.

Lambert, MJ.; Barley, DE. Research summary on the therapeutic relationship and psychotherapy outcome. In: Norcross, JC., editor. Psychotherapy relationships that work. Oxford University Press; New York: 2002. p. 17-32.

Pascual-Leone A, Greenberg LS, Pascual-Leone J. Developments in task analysis: New methods to study change. Psychotherapy Research. 2009; 19:527–542. [PubMed: 20183405]

Rice, LN.; Greenberg, LS. The new research paradigm. In: Rice, LN.; Greenberg, LS., editors. Patterns of change: Intensive analysis of psychotherapy process. The Guilford Press; New York: 1984. p. 7-24.

Tang TZ, DeRubeis RJ. Sudden gains and critical sessions in cognitive-behavioral therapy for depression. Journal of Consulting and Clinical Psychology. 1999; 67:894–904. [PubMed: 10596511]

Tang TZ, DeRubeis RJ, Beberman R, Pham T. Cognitive changes, critical sessions, and sudden gains in cognitive-behavioral therapy for depression. Journal of Consulting and Clinical Psychology. 2005; 73:168–172. [PubMed: 15709844]

Tang TZ, DeRubeis RJ, Hollon SD, Amsterdam JA, Shelton RC. Sudden gains in cognitive therapy of depression and depression relapse/recurrence. Journal of Consulting and Clinical Psychology. 2007; 75:404–408. [PubMed: 17563157]

Thase ME. How should efficacy be evaluated in randomized clinical trials of treatments for depression? The Journal of Clinical Psychiatry. 1999; 60:23–31. [PubMed: 10086480]

Wiles N, Thomas L, Anbel A, Ridgway N, Turner N, Campbell J, Garland A, Hollinghurst S, Jerron B, Kessler D, Kuyken W, Morrison J, Turner K, Williams C, Peters T, Lewis G. Cognitive behavioural therapy as an adjunct to pharmacotherapy for primary care based patients with treatment resistant depression: results of the CoBalt randomized controlled trial. The Lancet. 2012; 381:375–384.
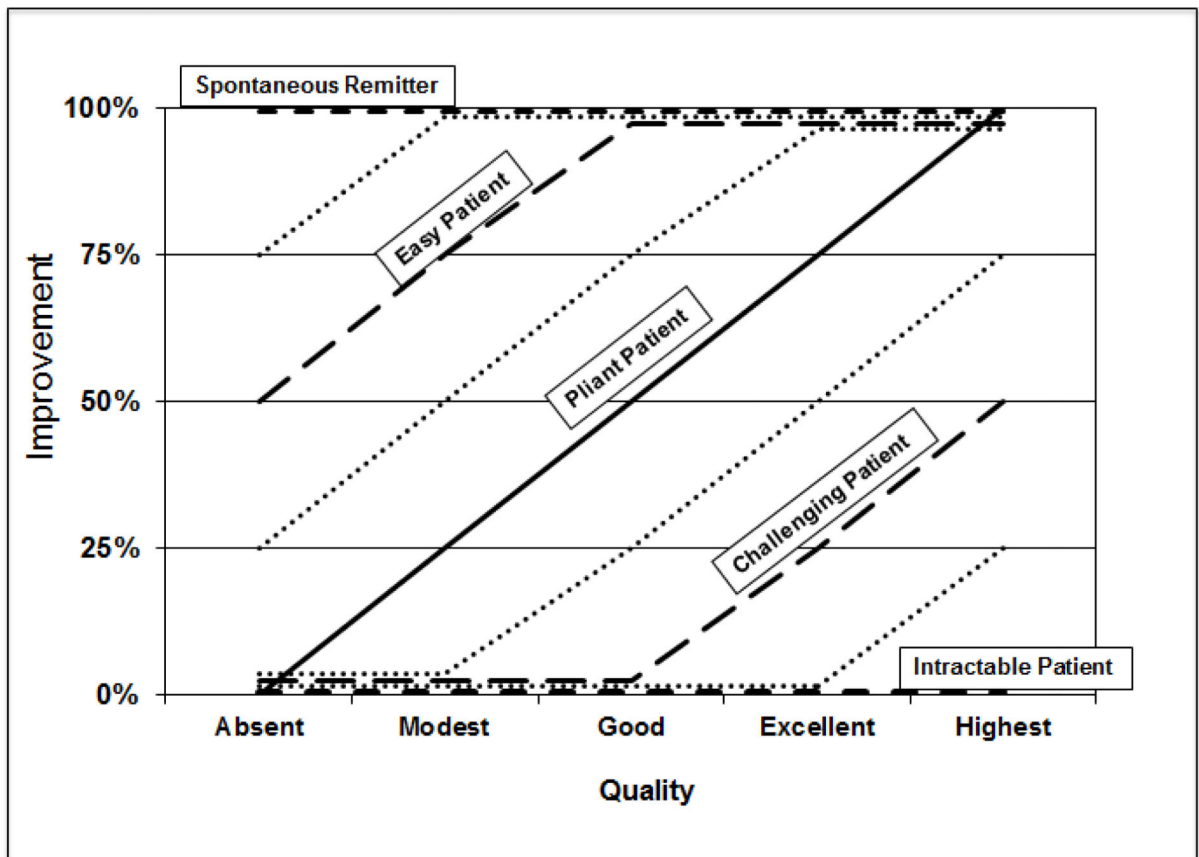
Figure 1.
Relationship between Therapy Quality and Improvement for Five Named Response Patterns,
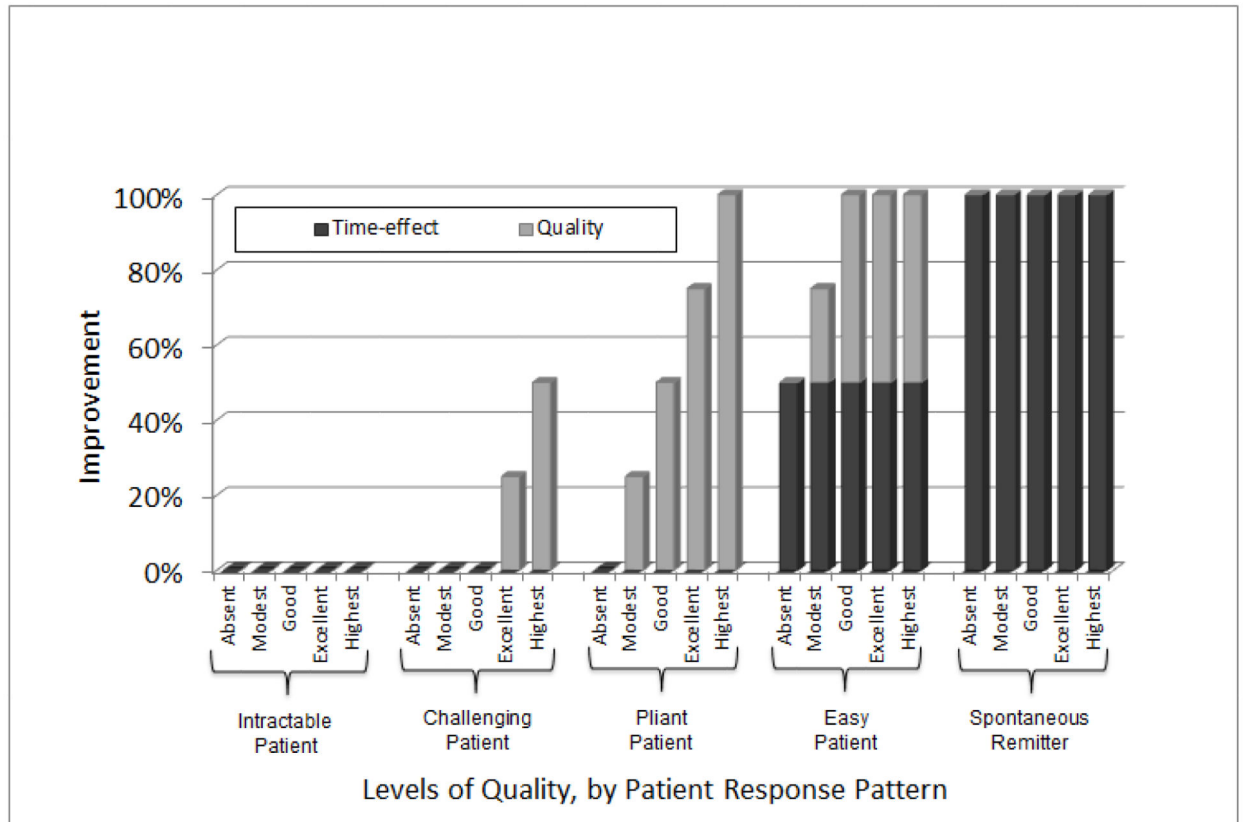with Four Intermediate Patterns.

Figure 2.
Example of the Effects of Time and Quality on Improvement in Five Hypothetical Patients.
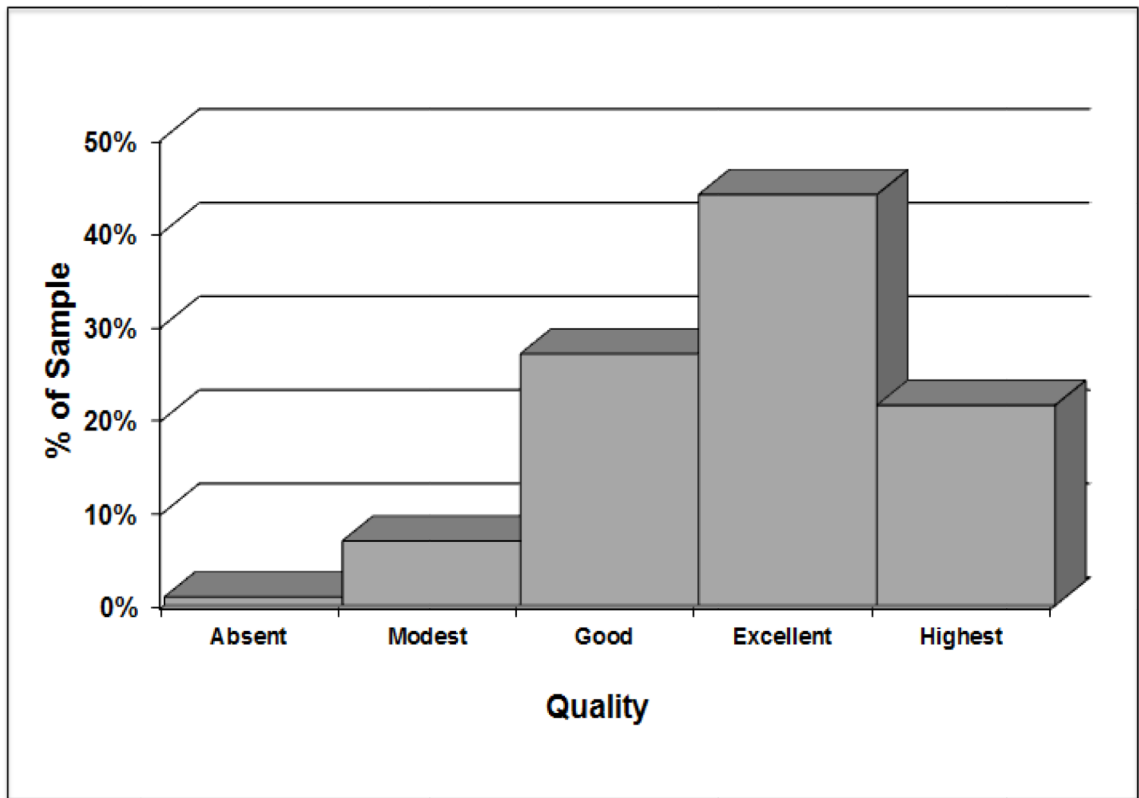
Figure 3.
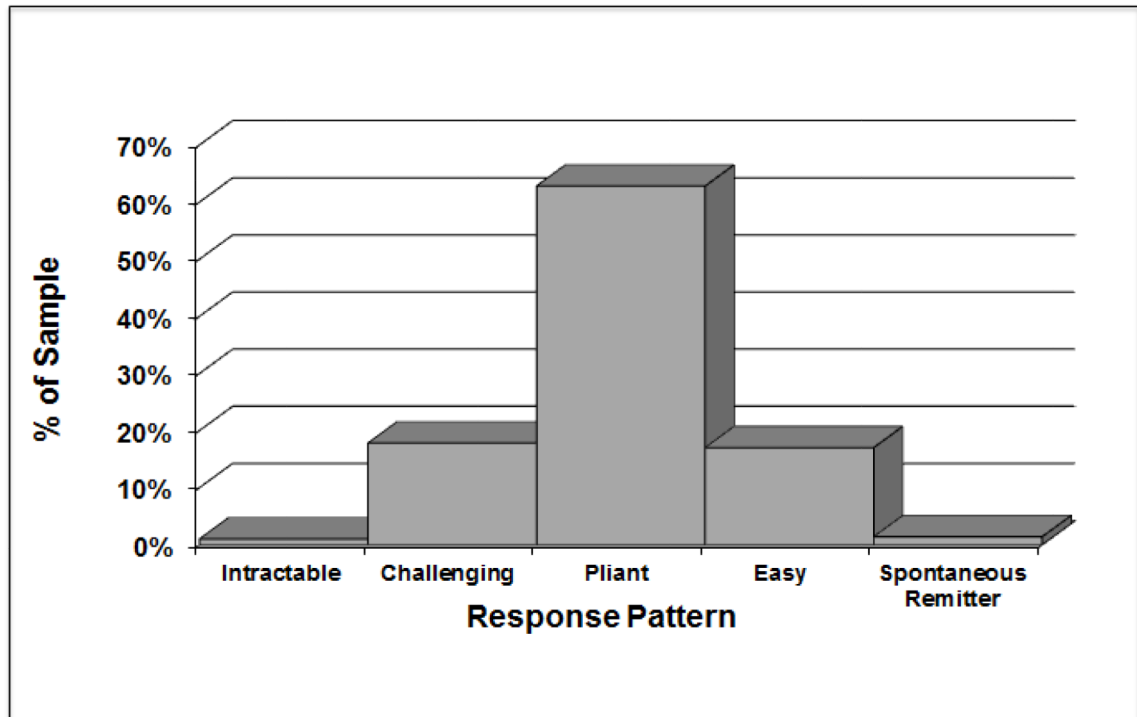A Skewed Distribution of Therapy Quality in a Study Relating Quality to Improvement.

Figure 4.
A Distribution of Response Patterns that will enhance the Likelihood of Identifying Therapy Quality Factors that Maximize Improvement (Predominantly Pliant).

Table 1

Correlations between Quality and Improvement from Simulations that represent Quality and Response Patterns with a Variety of Distributions

Case	Distribution of Therapy Quality	Distribution of Response Patterns	Assumes perfect measurement validity for both *quality* and *improvement*	Assumes that the measurement validity of both *quality* and *improvement* = .80
1	Uniform	Uniform	.33	.21
2	Skewed	Uniform	.18	.12
3	Skewed	Skewed	.39	.25
4	Skewed	Predominantly Pliant	.59	.38
5	Uniform	Predominantly Pliant	.79	.51