

## **Beyond a single standard: levels of evidence approach for evaluating marriage and family therapy research and practice**

Thomas L. Sexton,<sup>a</sup> Jeremy C. Kinser<sup>b</sup> and Christopher W. Hanes<sup>c</sup>

Randomized clinical trial (RCT) research has come to dominate the research landscape of marriage and family therapy (MFT). Despite becoming the 'gold standard' for evaluating clinical research and clinical practices, there is a growing debate regarding the reliance on RCTs as the primary basis for evaluating clinical intervention in MFT. Given the natural diversity of clients, settings and clinical problems faced by practitioners and the relational and recursive interactional process of MFT, one of the major challenges for the field of MFT will be to come to grips with the research–practice gap by moving beyond a single methodological standard through adopting a 'levels of evidence' approach as a framework that promotes diverse research methods, different methodological criteria (depending on the method), and evaluation based on the accumulated type of evidence needed to answer a specific policy, clinical practice choice, or within a model clinical decision.

### **Introduction**

In the current era of accountability, verifying clinical methods by research evidence through random control trials (RCT) has come to dominate the research landscape of marriage and family therapy. These methods promise to provide reliable and valid evidence for the efficacy of clinical interventions by reducing the variation in clients, contexts, therapists and interventions. Consequently, this approach produces evidence that the intervention or treatment model under consideration is responsible for the successful outcomes. This is an approach that is well established in the medical field where RCTs are the 'gold standard' for studies used to validate and determine what drugs are safe and reasonable to use in treating diseases. In the

---

<sup>a,b,c</sup> Center for Adolescent and Family Studies, Indiana University, 1901 East 10th Street, Bloomington, IN 47405-1006, USA.

United States, a number of professional organizations (e.g. Division 12 which is the Clinical Psychology Division of the American Psychological Association) have adopted RCTs as not just a standard of research but as the standard by which 'validated' and acknowledged practices are judged. In Great Britain there is a growing tradition of using research evidence, most frequently from RCTs, as the basis of both medical and behavioural health treatment decisions. For example, The National Institute for Health and Clinical Excellence (NICE), a part of the National Health Service, was formed to synthesize and establish best practices guidelines from published literature. NICE makes use of review-level literature, evaluation of individual RCTs and in particular meta-analyses, giving the highest weighting to RCTs which are assumed to have the greatest methodological rigour and thus a better standard. As a result, the RCT approach is becoming the 'gold standard' for evaluating clinical research and clinical practice.

There is, however, a growing debate regarding the reliance on RCTs as the primary basis for evaluating clinical interventions in MFT. Given the important implications for clinical practice, training and supervision, and the need for accountability and systematic inquiry, these debates often polarize the profession. This polarization results in an ever-increasing gap between the knowledge developed in research and the clinical practice of marriage and family therapy. Addressing this gap is important because it has the potential to stand in the way of bringing the most reliable, current and valid information to the treatment of those clients who seek our help.

The current reliance on this single method (RCTs) is understandable in two ways. First, historically MFT needed to substantiate its value by establishing itself as an effective practice relative to other treatments (Alexander *et al.*, 2002; Lebow and Gurman, 1995; Sexton *et al.*, 2004). MFT found legitimacy by adopting RCTs, methods held in high esteem by other clinical disciplines. These methods do provide a systematic process for gathering evidence and there is no doubt that the RCT is a powerful and useful research tool. RCTs can and do produce a valuable type of evidence that is useful for identifying and verifying the efficacy of clinical interventions. However, given the current developmental state of MFT practice and research the most relevant clinical questions go beyond what can be determined by any single method. The complexity of client presenting problems, the diversity of clients in regard to race, ethnicity and culture, the variation due to models' specific factors, common therapeutic factors and therapist variables all produce a set of interactions that are

complex (Sexton, 2007; Sprenkle and Blow, 2007). In fact, as our knowledge of the outcome of MFT has improved, RCTs seem to be a somewhat 'blunt' method (Eisler, 2007) in that they no longer produce the specificity of evidence needed for many of the clinical decisions surrounding the use and application of clinical interventions. In addition, concern over the potential limitations of RCTs as a single model has had the unintended consequences of furthering the existing research–practice gap in the profession. For many practitioners, RCTs are synonymous with research in general and concerns about this one method (RCTs) become confused with research in general. Thus, the already existing gap between the research and clinical perspectives is intensified and further widened (Westen *et al.*, 2004).

We suggest that one of the major challenges for the field of MFT will be to come to grips with the research–practice gap by moving beyond a single methodological standard and a narrow definition of knowledge. This is fostered by the adoption of a methodologically diverse approach to research and the resultant criteria for evaluating clinical practices. At the same time, we suggest that the field must retain the highest of methodological rigour by applying the best research methods to the types of clinical questions at hand. This allows one to determine the most useful clinical practices with varied clients, in varied settings, and with varied problems. We think this can be done without sacrificing the reliability and validity of good science by adopting a 'levels of evidence' approach as a framework to help integrate critical questions with the appropriate scientific method. A 'levels of evidence' model promotes diverse research methods, different methodological criteria and evaluation based on the accumulated type of evidence needed to answer a specific policy, clinical practice choice, or within-model clinical question. This approach would promote not just efficacy and effectiveness research but the additional use of multiple case studies, quasi-experimental work and research to provide ever increasing ecological validity and clinical relevance to clinical practice. It further promotes the clinicians, policy-holders and funders to be good consumers of research by matching their question of interest with the appropriate specificity of research evidence based on diverse methods and standards. When research is used in this way we suggest it will be an increasingly accepted and clinically relevant way to determine the best practices, guide therapist clinical decision-making, and promote systematic policy decisions.

Our goal in this paper is to make the case for a methodological diversity in research, well-articulated clinical techniques and

treatment programmes as the foundation of evidence-based practices. To accomplish this task we first consider issues in the debate regarding the current RCT gold standard and identify ways in which research and practice may benefit from an expanded view. Second, we describe a 'levels of evidence' alternative and identify what such an approach may require. Finally, we attempt to delineate what it will take to instate a 'levels of evidence' approach in the MFT field. The ultimate goal is to suggest an alternative model for using and conducting MFT research that goes beyond a single standard and method in order to capture the diversity of practice, clients and settings of MFT.

### **Randomized clinical trials: the 'gold standard' of clinical research?**

RCTs are rigorous and powerful research methods that provide a critical tool to the determination and identification of 'what works'. The 'power' of these approaches lies in the randomization of participants, and control group designs that optimize internal validity by minimizing and eliminating error allowing for the possibility of reliable and valid causal claims regarding efficacy and effectiveness of clinical practices (Kazdin, 2006). These are important methods for questions including 'what works' for particular problems, populations and within certain contexts. RCTs have also contributed largely to the establishment and promotion of evidence-based treatments and have become a 'gold standard' upon which clinical interventions are evaluated (Kazdin, 2006; Wampold and Bhati, 2004). In these ways, RCT-based research has led to improved services for a range of clients.

While RCTs play a key role in MFT research and the evaluation of clinical practice, they are not without their weaknesses, particularly given the current developmental state of the MFT field. In fact, the very strengths of the RCT approach may be its inherent weakness. In the early stages of the field, determining what worked was an appropriate level of specificity given that the practices had not been tested. Almost three decades later, the more important questions may be those that are focused on the contextual and ecological validity of models and techniques and clinical change mechanisms that operate to result in successful outcomes. Thus, it may be that RCTs with their emphasis on minimizing error variance also reduce the fundamental complexity of clinical practice. In the real world of MFT practice, clients and clinical problems are by definition diverse and complex.

Current clinical research does not always address the accompanying complexity of practice; therefore, there are increasing calls to consider multiple perspectives, cultural, ethnic, racial and gender diversity as a rule rather than as an exception (Westen *et al.*, 2004). Thus, in the current state of MFT practice it may be the moderators and mediators that influence the relationship between treatment and outcome that are the more clinically important questions. The diversity of moderating factors including service delivery systems as well as community values and cultures are critical in understanding outcomes.

For example, in order to minimize error in RCTs in psychotherapy research, client populations are often restricted to a limited range of diagnoses. Consequently, the interaction of problems and the diversity of clients are limited due to a high exclusion rate of the population studied (Messer, 2004). Thus, while providing internal consistency, the variation expected as an inherent part of clinical practice in community settings is reduced rather than embraced. RCT methodology also relies on treatment comparisons based on randomizations and 'treatment as usual' or quasi placebo controls that may have similar limitations. In medical research, placebo controls are rather simple to establish through double-blind procedures in which the active ingredient of the intervention is hidden from the participant. Given the complex relational and interpersonal nature of psychotherapy removing the 'active ingredients' is more difficult. Wampold and Bhati (2004), among others, questioned whether a 'no-treatment' or quasi placebo control can exist in psychological research. Wampold and Bhati (2004) argued that in psychotherapy research the active treatment is distinguishable from placebo treatment; 'treatment-as-usual' conditions rarely involve 'no treatment'. Therefore, these studies lack the same experimental validity found in medical research designs.

Westen *et al.* (2004) recently provided a comprehensive overview of the problems inherent in using RCT-based research as the primary basis of validating and identifying reliable clinical treatments. Practice guidelines based primarily on RCT standards of research overstate the value of findings, particularly with regard to the complexities of clinical practice. Westen and colleagues' concerns are levelled not at the role of research in clinical practice, but at the implicit assumptions of the existing guidelines. They suggested that empirical support is not a decision made between a limited set of 'lists' but instead requires many types of evidence to account for the complexities of clients, therapists and treatment settings in evaluating the research. In

addition, they argued that the complexities confronted in clinical practice require a more 'nuanced' view of treatment outcome that addresses the co-morbidity of clients, and necessitates a focus on problems and/or profiles of clinical problems that extend beyond the categories of the restrictive and often ecologically invalid DSM.

It is important to note that RCTs and the methodological issue related to these research approaches do not only apply to individual clinical trial research studies. Over the past number of years, meta-analytic techniques have been applied as a means to help understand the value of psychological interventions (Wampold, 2001). Meta-analysis is a powerful tool for aggregating across studies to overcome the effects of small sample sizes. Most of these studies also include a metric to judge research quality that is used to adjust the effect size. RCT studies are most commonly given great weight, and thus great impact on the findings of such studies. Thus, criticism of RCT designs is also appropriate when considering that meta-analytic review is becoming increasingly common as a measure of the impact of clinical effectiveness.

### **A level of evidence approach: integrating research and practice**

We are not suggesting that the field abandon RCTs as an important research approach and clinical outcome standard. Instead we suggest that RCTs are a necessary but not sufficient approach to understanding, evaluating and promoting effective practices in MFT. MFT treatments may be regarded as having various levels of evidence from the broad (does it work compared to no treatment) to the specific and clinically nuanced (why does this work in this situation with this person). Determining what are 'good' treatments would be based on these different methods matched to the 'level' of evidence most appropriate. For example, in determining what works, RCTs provide a valuable tool to validate absolute and relative efficacy (Kazdin, 2006). However, once established, alternative approaches are necessary to answer the more 'fine-tuned' and clinically rich questions. These methods may include case studies, matched control designs and meta-analysis. To do so, each method would need to be used when it fits the question at issue rather than used for its exclusive value while at the same time meeting the established methodological quality that fits the method at hand.

This approach is based on the assumption that there are different types of evidence that help guide practice. For instance, absolute

effectiveness evidence is a measure of the success of the intervention compared to no treatment. Such a comparison is useful in determining if an intervention may be considered evidence-based. Relative efficacy is the comparison of an intervention to a reasonable alternative (e.g. common factors, a treatment of a different modality, or a different intervention). Relative efficacy is critical to establish that a treatment is the best choice for a specific client/problem. However, when more fine-tuned clinical questions are of interest, Sexton *et al.* (2004) suggested the importance of contextual efficacy, defined as the degree to which an intervention is effective in varying community contexts. This provides a critical dimension and a final level that focuses on change mechanisms within particular practices.

The concept is best illustrated in the model proposed by Sexton *et al.* (2004) as they developed a conceptual framework for their comprehensive review of the MFT literature. They argued that diverse research methodologies offer unique perspectives from which to judge the evidence of an intervention or treatment programme. With different questions and in different developmental contexts the most appropriate research method is likely to change. For example, RCTs emphasize high levels of internal validity and are critical to establishing the evidence for an intervention to be considered evidence-based. *Outcome studies* investigate the absolute (as compared to no treatment) and relative efficacy (as compared to a clinically legitimate alternative intervention) of an intervention or treatment programme. *Comparison trial* studies compare a couple or family intervention to a systematically developed and relevant comparison intervention or treatment. The value of each type is determined by the degree to which it is the most appropriate method to answer the question at hand. The focus is on differences in clinically significant outcomes that represent client improvement. *Efficacy studies* answer questions about which treatments work under the most stringently controlled conditions. *Effectiveness studies* answer questions regarding the power of therapeutic interventions in actual clinical settings with conditions that replicate those that actual clinicians face. Although there is decreased methodological control (in the traditional sense), these studies have high clinical relevance. Effectiveness studies are often conducted in community settings where the experimental control of traditional efficacy studies is not possible. *Moderator studies* find the degree to which a certain client, problem or context feature may moderate the existing outcomes. *Process-to-outcome* studies link the conditions of therapy (pre-existing and specific within-session

processes) with the outcomes of family-based interventions. These studies help identify the mechanisms of action in an evidence-based intervention/programme. *Systematic case studies* provide an ideographic view of the clinical process. These studies are particularly useful in identifying the individual experiences in the change process that might lead to a better understanding of clinical mechanisms or outcomes. *Transportability* studies consider various issues related to the transportation of MFT interventions/treatments to the community settings where they might be practised. Such studies might consider the contextual variables (e.g. therapist variables, client variables, organizational service delivery systems) that may either enhance or limit successful community implementation. *Qualitative and meta-analytic research reviews* help contribute to understanding and identifying: 'common' elements, new treatment mechanisms or differential results across studies (Edwards *et al.*, 2004).

Adopting a diversity of methods perspective must be accompanied by high standards of methodological rigour. Given the diverse methods, there can be no single standard of methodological excellence. Instead, the standard used to evaluate evidence must match the type of study. In order to be the basis of relevant clinical intervention, high-quality studies of family psychology interventions/treatments should include clear specifications regarding the contents of the intervention/treatment model (e.g. manual), measures of intervention/model fidelity (therapist adherence or competence), clear identification of client problems, complete descriptions of service delivery contexts in which the intervention/treatment is tested, and the use of specific and accepted measures of clinical outcomes.

The value in the 'levels of evidence' perspective is in how it may be applied in answering clinical questions and ultimately in understanding the potential value of clinical interventions. Instead of applying a single perspective (e.g. RCT), a 'levels-of-evidence' approach suggests that there are varied types of questions to be answered by various constituents in the MFT field. To successfully apply the knowledge from diverse research methods, the question must be matched to the type and 'level of evidence' in order to make useful clinical decisions. For example, policy-makers ask about what practices or treatment models to use, how to spend money, and what practices to promote that address the pressing problems of the people and communities. These questions are probably best informed by broader 'levels of evidence' that address absolute and relative efficacy/effectiveness and transportation-based studies. For agency administrators, similar



service delivery questions arise regarding choice. For example, asking what practices to use, as well as evaluating the need to determine effectiveness in their local setting and determining treatment quality. These questions are probably best answered by model/treatment studies of the moderators of clinical outcomes that involve much more of a context type of efficacy (does it work in this situation given these complexities, can it be transported to this agency?). Clinicians need to know what to do 'in the room' and thus need a very fine-tuned and specific level of evidence focused on the change mechanisms and their interaction with differences in client and setting. These questions are best answered by change mechanism and process-based research. The knowledge from diverse methods applied to the correct clinical questions provides the most comprehensive view of good MFT treatment.

### **Levels of evidence in MFT: the Family Psychology Task Force (USA) on Evidence-based Treatments**

The Division 43 (Family Psychology) Task Force on Evidence-based Treatments in Family Psychology in the USA was established to develop and implement an approach to the identification of evidence-based clinical interventions in family psychology so that clinically useful treatments could be identified and made available to the public. The Task Force was a group of researchers, practitioners and trainers. The Task Force constructed guidelines that were appreciative of the need to attend to both the artfulness and individuality of effective clinical work *and* the invaluable role of research at all levels of clinical decision-making. A 'levels-of-evidence' approach was the basis of the guidelines (see Sexton *et al.* (2007) for a comprehensive discussion of the assumptions, principles and proposed uses of these guidelines). What follows is a brief outline of the 'level' of MFT clinical interventions when various levels of research evidence are considered. The model includes three levels (promising, evidence-informed and evidence-based). Within those models that are evidence-based, the four additional categories that describe increasingly specific and fine-tuned research knowledge are articulated (category 1: absolute efficacy, category 2: relative efficacy, category 3: change mechanisms, category 4: contextual efficacy). Once established with credible outcomes, evidence programmes (meeting the criteria of category 1) would continue to develop a further research base including systematic study of change mechanisms and studies of the contextual efficacy

and application of the model. The additional categories define evidence above and beyond these minimal levels, providing a systematic way to identify the evidence-based strengths and weakness of a treatment programme.

*Level 1: Evidence-informed interventions/treatments*

These are informed by previous research/basic psychological research, or a common factors perspective. The intervention/treatment programme uses interventions explicitly linked to the evidence or to portions of an evidence-based programme to suggest that they have an evidence base. Lack of evidence may be due to a less well-defined and articulate intervention/programme, non-specific clinical problems (and appropriate outcomes), or a dearth of research on the programme.

*Level 2: Promising interventions/treatments*

These are specific clinical interventions that have either preliminary results, evaluation outcomes, or only comparison level studies of high quality but no further evidence or specific outcomes with specific populations.

*Level 3: Evidence-based treatments*

These are those specific and comprehensive treatment intervention programmes that have systematic evidence that they work with the clinical problems they are designed to impact upon.

*Category 1: Absolute efficacy/effectiveness evidence.* This shows that the specific treatment intervention programme produces reliably improved clinically relevant outcomes when compared to the typical improvement rates for given clinical problems. These would be efficacy studies with comparison or clinical trial evidence that show clinically significant effects with specific clinical outcomes that have clinical relevance.

*Category 2: Relative efficacy/effectiveness.* This shows that the specific treatment intervention programme produces reliable and improved clinically relevant outcomes when compared to an alternative/viable treatment. This is a more difficult test than category 1 for demonstrating that the programme works because it requires gains beyond

those other treatments. This level of evidence suggests that this intervention/treatment programme is a clinically reliable treatment for a specific class of clinical problems.

*Category 3: Effective models with verified mechanisms.* These show evidence that the model-specific change mechanism operating within the intervention programme is linked to relevant identifiable outcomes, as theoretically expected. This level of evidence suggests that the treatment programme is a clinically reliable treatment programme for a class of clinical problems that operates through the described mechanisms to produce the demonstrated outcomes.

*Category 4: Effective models with contextual efficacy.* These show evidence that in addition to being effective (category 1) the model has successful outcomes (absolute and relative) with a range of clients, clinical problems and service delivery contexts. This evidence would suggest that the treatment programme is effective in some ranges of clinical contexts and is potentially transportable to those specific contexts, clients or problems. This level of evidence suggests that the programme produces change and that the outcomes are effective for specific client populations (i.e. gender, age, race, culture), clinical problems (i.e. behaviour disorders, depression, school problems) in specific service delivery make up the context within which the programme must work. This level of evidence would suggest that the treatment programme is a clinically reliable treatment programme for a class of clinical problems that is widely applicable.

## **Conclusion**

The seemingly ever-present gap in the field of MFT between research and practice has become more polarized with the advent of accountability and the emergence of research-based standards for evidence-based practices. This is an understandable and logical next step in the developmental path of the MFT profession. This has in part been fuelled by the adoption of a single research methodology (RCT) and the reliance and weighting of RCTs in meta-analytic studies. We suggest that to successfully move research and practice together, a dialectical approach is needed (Sexton *et al.*, 2004). This is one in which research and practice are viewed and experienced as just different sides of the same coin, both with the common goal of

improving practice to better help the diversity of clients and settings in which MFT is practised.

We suggest that to accomplish this goal, research should play a major role in studying and determining good practice. However, it must be research that is methodologically diverse so that a range of perspectives and levels of knowledge can be accumulated and brought to bear on the myriad clinical questions facing clinicians, administrators, policy-holders and clients. A levels-of-evidence approach may be used as a way to create a system that helps move valuable and relevant research evidence into clinical practice by creating a framework to systematically evaluate the available research in a manner that can efficiently be used to guide practice decisions while at the same time providing researchers with a way to identify clinically relevant and timely research about new or existing clinical approaches.

## References

- Alexander, J. F., Sexton, T. L. and Robbins, M. A. (2002) The developmental status of family therapy in family psychology intervention science. In H. A. Liddle, *Family Psychology Intervention Science*. Washington, DC: American Psychological Association Press.
- Edwards, D. J. A., Dattilio, F. M. and Bromley, D. B. (2004) Developing evidence-based practice: the role of case-based research. *Professional Psychology: Research and Practice*, **35**: 589–597.
- Eisler, I. (2007) Treatment models, brand names, acronyms and evidence-based practice. *Journal of Family Therapy*, **29**: 183–185.
- Kazdin, A. E. (2006) Arbitrary metrics: implications for identifying evidence-based treatments. *American Psychologist*, **61**: 42–49.
- Lebow, J. L. and Gurman, A. S. (1995) Research assessing couple and family therapy. *Annual Review of Psychology*, **46**: 27–57.
- Messer, S. B. (2004) Evidence-based practice: beyond empirically supported treatments. *Professional Psychology: Research and Practice*, **35**: 580–588.
- Sexton, T. L. (2007) The therapist as a moderator and mediator in successful therapeutic change. *Journal of Family Therapy*, **29**: 104–108.
- Sexton, T. L., Alexander, J. F. and Mease, A. L. (2004) Levels of evidence for the models and mechanisms of therapeutic change in family and couple therapy. In M. J. Lambert (ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (5th edn). New York: Wiley.
- Sexton, T. L., Coop-Gordon, Gurman, A. S., Lebow, J. L., Holtzworth-Munroe, A. and Johnson, S. (2007) Evidence-based treatments in couple and family psychology. *Report of the Task Force on Evidence-based Treatments in Family Psychology*. Division 43, Family Psychology American Psychological Association.
- Sprengle, D. H. and Blow, A. J. (2007) The role of the therapist as the bridge between common factors and therapeutic change: more complex than congruency with a worldview. *Journal of Family Therapy*, **29**: 109–113.

- Wampold, B. E. (2001) *The Great Psychotherapy Debate*. Mahway, NJ: Lawrence Erlbaum.
- Wampold, B. E. and Bhati, K. S. (2004) Attending to the omissions: a historical examination of the evidence-based practice movement. *Professional Psychology: Research and Practice*, **35**: 563–570.
- Westen, D., Novotny, C. M. and Thompson-Brenner, H. (2004) The empirical status of empirically supported psychotherapies: assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin*, **130**: 631–663.