

An overview of research methodology in the development of family-focused treatment programs

Paul Harnett & Sharon Dawe

BIOGRAPHY

Dr Paul Harnett PhD is a clinical psychologist who has worked in the areas of child protection and juvenile justice for 15 years, both in England and Australia. He has recently been involved in research investigating the effectiveness of a prevention of depression program targeting adolescents.

Dr Sharon Dawe PhD has worked in the area of substance abuse for 15 years both as a clinician and a researcher.

Paul and Sharon are currently evaluating the effectiveness of the Parents Under Pressure (PUP) program in families who are on methadone maintenance.

Introduction

As practitioners working with families, we like to think that our clinical work has benefits for our client group. However, it is probably fair to say that practitioners vary in the effort they put into providing objective evidence that their interventions are achieving the intended benefits for their client group. In an age of accountability, agencies are increasingly insisting that evidence is produced, through the ongoing evaluation of clinical practice, to demonstrate that the expense of providing services is justified. Further, funding agencies in health, welfare, education and the justice systems are adopting evidence-based models of service delivery.

This means that practitioners are being encouraged to deliver interventions¹ that have been shown to be effective in clinical research trials. The assumption is that interventions with empirical support of their effectiveness will be more beneficial to the client group the agency serves than ad hoc or untested interventions.

For practitioners providing services to fathers, the emphasis on evidence-based practice creates some difficulty, given the limited range of interventions targeting fathers that have been subject to empirical test. Practitioners will be keen to see the interventions that are being developed (as described in this and other volumes such as Dudley & Stone, 2001; Fagan & Hawkins, 2001) are subject to empirical investigation. However, even after clinical trials have demonstrated the positive impact of an intervention, it cannot be assumed, for reasons discussed later in the chapter, that the intervention will be effective with a particular father in a specific clinical setting. The response of an individual to an intervention will always require individualised evaluation.

In this chapter we discuss a number of issues concerning the evaluation of services. Our aim is to provide information that practitioners can use. First, we present an overview of research methodology to provide practitioners with basic knowledge to critically appraise the empirical evidence supporting available treatments (Gaynor, Baird & Nelson Gray, 1999; Shapiro, 1967, 1985).

Second, we provide an introduction to the methodology appropriate for practitioners to evaluate the effectiveness of interventions in their own clinical practice, that is, single-case methodology (Barker, Pistrang & Elliott, 1994; Blampied, 2001; Hayes, Barlow & Nelson Gray, 1999; O'Gorman, 2001; Raimy, 1950; Stricker & Trierweiler, 1995; Trierweiler & Stricker, 1998; Turpin, 2001).

To illustrate the issues, we present an overview of the methodology adopted in the research evaluation of the Parents Under Pressure (PUP) program, a parenting program recently developed that targets the needs of multi-problem high-risk parents (Dawe, Harnett & Rendalls, 2000).

The two major approaches to the evaluation of clinical treatments are: single case methodology involving the evaluation of change across time

1 The term 'intervention' is used to refer to prevention, early intervention or treatment programs.

in an individual case; and group comparisons in which a new treatment is compared to standard care or current practice. Both forms of evaluation aim to demonstrate the internal validity of an intervention. In other words, they aim to show that the observed difference between participants who received and those who did not receive an intervention is due to the effect of the intervention and not to unrelated variables or errors in measurement (Blampied, 2001; Kendall, Butcher & Holmbeck, 1999; Trierweiler & Stricker, 1998). These are discussed in turn below.

Single case designs

Single case methodology aims to demonstrate the internal validity of an intervention, that is to determine whether behaviour change is attributable to the intervention and not to unrelated, extraneous variables (Blampied, 2001; Gaynor et al., 1999; Kazdin, 1981). The essence of single-case methodology is the analysis of data derived from repeated measurement of a client's behaviour over time. This involves the independent variable (the intervention) being systematically varied and related to change in the dependent variable (the target behaviour).

Support for the internal validity of the intervention is found when variation in the intervention is systematically related to changes in the client's behaviour (Blampied, 2001).

There are several commonly used single-case designs that may be used with individual cases. As it is beyond the scope of this chapter to describe these, one example, a withdrawal design, will be presented to illustrate the logic of single case methodology.

Withdrawal designs involve repeatedly withdrawing and reintroducing an intervention. In this situation it would be expected that the client's functioning should deteriorate and improve as the intervention is withdrawn and introduced respectively.

For example, a parent may be prompted in weekly meetings to ignore a child's whining behaviour that has been leading to tantrums. Measures of the frequency of tantrums show that the intervention was associated with a decrease in the frequency of tantrums. However, it cannot be ruled out that the child's increased maturity or some other event caused the tantrums to subside. The therapist stops prompting the parent to ignore

the child's whining and this is followed by an increase in tantrums. Finally, the therapist reintroduces the prompting to ignore whining, and the tantrums again decrease in frequency. Replicating the impact of the therapist's prompts in this situation provides strong evidence that ignoring the child's whining was the active ingredient in decreasing the tantrums.

Unfortunately, it is often not possible or ethical to withdraw or withhold an intervention in clinical practice. Further, a permanent change in behaviour may occur that is not reversed through the withdrawal of an intervention. Alternative single case designs have been developed to address such problems (Gaynor et al., 1999; Hayes et al., 1999).

Group designs

Many clinical research trials use group designs where groups of individuals receive either treatment A or treatment B. By randomly assigning individuals to a group, the likelihood is that the groups will be similar on unmeasured factors that could possibly be related to the intervention.

As an example, consider a parenting program to help fathers manage the behaviour of their children. One factor that could limit the impact of the program is the level of behaviour problems displayed by the children. Random allocation attempts to ensure the level of conduct problems is not greater in one group compared to the other. Thus, there is no reason to expect that one group would change any more or any less than the other.

In group designs it is expected, however, that participants in the study will change over the course of the study. Factors such as maturation, unmeasured environmental factors and testing effects (Mitchell & Jolley, 1996) can all result in some change.

For example, a program for fathers who have recently separated from their partners may find that a number of the fathers enter new relationships over the course of the intervention, which could influence their functioning in the father role. However, these changes are expected to occur equally *within* the two groups. If this is the case, the changes in each group that are due to the influence of fathers entering new relationships will be the same for each group. On the other hand, if an intervention were

delivered to only one group, it would be expected that there would be measurable differences between the groups following the intervention. Thus, any changes *between* the groups that do occur are attributed to the effects of the intervention (Kendall et al., 1999).

This design, referred to as a randomised controlled trial (RCT), is a powerful methodology for demonstrating the impact of an intervention. However, inadequate reporting of the design, conduct, analysis, and interpretation of a RCT can provide misleading information about the effectiveness of the intervention that has been tested (Moher, Schulz & Altman, 2001).

The CONSORT statement (Begg, Cho, Eastwood, Horton, Moher, Olkin et al., 1996) was developed by an international group of clinical researchers, statisticians, epidemiologists, and biomedical editors to standardise and improve the quality of published reports of RCTs. The CONSORT statement provides a useful means for practitioners to assess the strengths and limitations of a RCT. Readers interested in a detailed discussion are referred to the CONSORT website: <http://www.consort-statement.org>. A summary of the key issues to consider when determining the methodological soundness of a RCT follows.

Comparison or control condition

The most methodologically robust RCT is one in which the intervention group and the control group are closely matched on all factors possibly related to outcome.

For example, if an intervention was expected to be more effective for, but not limited to, fathers in a current relationship, it would be necessary to ensure that the proportion of fathers in a current relationship was similar in the two groups. Similarly, if a study was comparing a five-session program with a ten-session program that included additional material, it would not be possible to determine whether it was the additional components of the program that resulted in the additional benefits, or whether it was simply that participants received more sessions.

Comparison groups therefore need to be similar. This raises a number of ethical issues in terms of offering treatment intervention versus not offering, or placebo treatment. There is no single solution to these problems and it is suggested that innovative interventions should be evaluated using a range of designs, preferably by independent research groups.

Randomisation

Having decided that the study should include a comparison condition, it is necessary to ensure that participants are randomly allocated to the treatments. Random allocation to different treatment conditions presumes that extraneous variables or factors will be equally represented in all groups involved in a study. Changes in participants resulting from these uncontrolled factors should then occur equally within the two groups. Randomisation allows the evaluation to highlight the differences between the groups that are due to the intervention, while reducing the influence of bias. Randomisation is probabilistic, so the larger the sample size, the more likely it is that the extraneous variables will be equally distributed, thereby minimising bias even more (Hsu, 1998). However, we appreciate that using this technique is not always possible in practice settings.

Where the use of random assignment is not possible, an alternative could be, for example, to allocate families in accordance with the level and intensity of the problems experienced and the intensity of the intervention required. It must be noted that the outcomes of this approach would be biased. The CONSORT guidelines describe procedures for the random allocation of participants to intervention conditions (Moher et al., 2001).

Treatment Integrity

'Treatment integrity' refers to the quality of program implementation. That is, the extent to which an intervention was implemented competently and according to the research protocol specified in the treatment plan or manual (Waltz, Addis, Koerner & Jacobson, 1993). Compared to most medical studies of a new drug regime or procedure, defining the actual parameters of a clinical or psychosocial intervention raises many difficulties.

Treatment programs often consist of multiple components. These components may be highly structured, delivered in a specified order with clear instructions about the terminology that should be used by the clinician. Good examples of highly structured, manualised interventions, are the project MATCH manuals used to determine the relative effectiveness of motivational interviewing, 12 step and relapse prevention in the treatment of alcohol dependence (NIAAA Project MATCH Monograph Series, 1994).

Other treatments may be less structured, and developed from a potential suite of strategies and interventions that are tailored specifically for individual cases or families. Whilst the use of highly structured manualised interventions allows for a precise assessment of treatment adherence, such approaches are not without their problems. There has been a growing acknowledgment that such approaches ignore the importance of the therapeutic process that occurs between client and clinician (Kendall & Chu, 2000). Indeed it may be that the therapeutic alliance is as important as treatment content in successful outcomes. Further, it has been argued that clinicians may be unwilling to adopt highly prescriptive treatments due to a reluctance to be limited by a prescribed treatment protocol (Clarke, 1995). However, interventions that are less rigidly manualised or set out will result in greater variation in the delivery of the intervention, including variation among practitioners and also by the same practitioner across time (Waltz et al., 1993).

Whether a treatment manual contains session-by-session outlines of interventions, or describes broad principles and phases of treatment, including examples of interventions, some check that the practitioner has delivered the treatment in accordance with the associated framework, procedures or instructions is essential. How this is done varies considerably from case to case and may include audio or videotaping of sessions, completion of therapist checklists by clinicians or treatment checklists completed by participants. The most appropriate method for a particular study will depend on the structure of the intervention, and practical considerations such as the impact of video cameras on the therapeutic process.

Potential bias in assessment and at follow up

Determining whether a new treatment has been of benefit to the participants requires that measures of baseline functioning of the participant are undertaken and that these measures are repeated across time, usually at the end of treatment and again at some later point. This helps determine some benchmarks for the treatment intervention.

The CONSORT guidelines propose that these assessments, especially at follow up, should be conducted by an independent researcher. This will help increase the objectivity of the evaluation and decrease the possibility of bias or practitioner influence on a participant's self-report. This

is particularly important in the evaluation of intensive family-based interventions as the clinician has often worked closely with the family over an extended time and it is possible that a family's responses would be influenced by their feelings towards the clinician.

Issues of measurement

In order to determine whether an intervention is associated with improved outcome it is necessary to articulate clearly what aspects of the presenting problems are most likely to change given the content of the intervention.

For example, researchers of a parenting program that focuses on behavioural management strategies for non-compliant children may be interested in determining whether the intervention is associated with decreases in a child's non-compliance or aggression. Or, researchers of a parenting program that focuses on extending parental social networks would be more interested in measuring the parent's perceived support in response to the program.

Measuring the conceptual domains targeted by the intervention

Given the complexity of family problems and the current focus on the ecological context in which families are embedded (Belsky, 1984; Cicchetti & Toth, 1998), parenting programs are increasingly aiming to effect change across multiple domains. These include changes in the psychological functioning of the parents, parent-child relationships, and the wider social context of the family (Dawe, Harnett, Staiger & Dadds, 2000; Sanders, 1995).

Within each domain, it is possible to further identify targets for change such as the psychological health of each individual within the family and the quality of the relationships between members of the family. Changes in the wider social environment of the family are also often targeted for intervention and may be measured by the use of external services and the extent by which the participants themselves perceive external agencies as providing support. Thus, researchers evaluating a parenting program need to define the nature of the changes they predict and ensure that the measures accurately reflect the proposed areas of change.

Reliability and validity

Chosen measures also need to be reliable and valid. The term 'reliability' refers to the consistency with which an instrument measures a particular construct, or aspect and is evaluated in several major ways.

'Test-retest reliability' refers to the extent to which an instrument provides a similar measure of a particular construct at two points in time. The interval between these two time points should be fairly brief so that the measurement is not affected by the passage of time. Also, no intervention should be administered during this period. If a measure is not shown to have test-retest reliability, fluctuations in scores may simply reflect measurement error.

Another form of reliability is internal consistency. 'Internal consistency' refers to the extent to which items within the method used (for example, questions of a self-report questionnaire) provide a similar measure of the construct or aspect being measured. Questionnaires, for example, generally include a number of items measuring the same construct. A questionnaire measuring stress in the parenting role may elicit responses to the following statements:

- I find it stressful to manage my responsibilities as a parent.
- Being a parent is more stressful than I had imagined.

Clearly the participant should rate the answers to these questions in a similar manner. If it was found that items allegedly measuring the same construct were not well correlated, the measure is said to have poor internal consistency.

The 'validity' of a method or instrument refers to the extent to which it accurately measures the construct it is believed to measure. There are different types of validity, including content validity and construct validity.

Content validity is where the method used measures all aspects of the particular construct being evaluated. For example, a measure of depression should include items that capture the components of lowered mood, sleep disturbance, and appetite change (Barker et al., 1994). A measure is said to lack content validity if it fails to measure important aspects of the construct under study.

Construct validity is the success of the final scale to measure the characteristics of the construct. This can be demonstrated in a number of

ways, of which one method is to show that scores on the measure are highly correlated with another measure of the same construct and uncorrelated with unrelated constructs. For instance, a measure of parental stress should correlate highly with a measure of anxiety and poorly with a measure of reading ability. A more detailed discussion of general issues in validity can be found in Anastasi and Urbina (1997).

Other issues of measurement

How constructs or aspects relevant to child and family psychology are measured falls into three broad categories:

- standardised interviews
- self-reports
- observational measures

In the first, a structured interview consisting of a series of questions is conducted by the interviewer who then uses a rating scale to provide a measure of the conceptual domain that is addressed by each question. This may be dichotomised as a yes/no response or rated on a scale. Such approaches are often used to determine diagnostic status within the current classification system for mental disorders: the *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition (American Psychiatric Association, 1994).

Self-report measures are typically less concerned with diagnosis as such and more with determining where an individual fits on a continuum. Some self-report instruments that are concerned with highly sensitive information that participants may be reluctant to report have validity checks. *The Child Abuse Potential Inventory* (Milner, 1986) is one such example. As parents may have occasion to either under-report or to overstate their child's problems this measure has both a 'faking good' and 'faking bad' scale in addition to other validity checks.

Finally, many studies in the child and family area use observational measures to determine whether there is an observable difference in families or children following an intervention. Ratings of the quality of parent-child transactions can focus on specific parent-child interactional sequences using measures such as the Dyadic Parent-Child Interaction Schedule (Eyberg & Robinson) or Family Observation Schedule (Dadds & Sanders, 1992). In such instances, behaviours are

coded according to strict criteria and the number of behaviours per criterion is counted to determine whether there has been a change. An alternative approach is to make global ratings of particular constructs such as parental sensitivity or hostility. One such observational system, the Emotional Availability Scales, has been developed by Biringen and colleagues (Pipp Siegel & Biringen, 1998).

Ideally, studies should include measures using different types of instruments. For example, a stronger case can be made that the quality of the parent-child relationship has improved if behavioural observations of the quality of the parent-child relationship is consistent with the parent's self-report of their relationship with the child (Roberts & Hope, 2001).

Distinguishing efficacy and effectiveness research

Methodology that focuses on the internal validity of an intervention is important in demonstrating what is possible under ideal conditions. Studies that address the internal validity of an intervention are referred to as 'efficacy studies' and almost always have strict inclusion and exclusion criteria for participants. The delivery of the intervention in efficacy studies is usually by selected and trained clinicians who receive ongoing supervision and adherence to the treatment protocol is ensured during supervision. Finally, the intervention in efficacy studies is delivered in places that are well resourced and designed to accommodate the particular intervention (such as university clinics, teaching hospitals).

Effectiveness studies, on the other hand, evaluate interventions under real-world conditions and address the external validity of the intervention (Chambless & Hollon, 1998). Effectiveness studies generally involve a more heterogeneous sample (that is, few or no inclusion and exclusion criteria), are conducted in 'real-world' settings, such as a home, school, general medical setting, and are implemented by the practitioners in the clinical setting.

Effectiveness studies, then, are characterised by less control over:

- the client population selected;
- the prior training, skills and motivation of practitioners implementing the program; and

- other setting events including the level of supervision available for practitioners, availability of time to prepare sessions, space in the clinic, and the commitment and support of colleagues and managers.

The aims and methodology of an effectiveness study vary from those of an efficacy study. Effectiveness studies are more concerned with the problems associated with the dissemination of an intervention. These problems include issues concerning the recruitment and engagement of families, the quality of program implementation when practitioners in clinical settings deliver the intervention, and the level of training and models of supervision required to support practitioners implementing the interventions. Organisational factors such as the restraints of time, space, and the support of management that can interfere with program delivery. Client outcomes and satisfaction with the intervention are also important (Clarke, 1995; Henggeler, Schoenwald, Liao, Letourneau & Edwards, 2002; Matthew R. Sanders, Turner & Markie Dadds, 2002; Weisz, Donenberg, Han & Weiss, 1995).

Development and evaluation of the Parents Under Pressure (PUP) Program

In this section the methodological issues raised above are illustrated by describing the development and research evaluation of the Parents Under Pressure program, a parenting program developed for multi-problem, high-risk parents (Dawe, Harnett & Rendalls, 2000).

The research literature on high-risk parents (including parents who abuse substances and those involved with child protection agencies) clearly indicates that in addition to parenting skills, these groups of parents have a number of areas of vulnerability that may impede their ability to protect and care for their child. Dawe et al. (2000) argued that a program targeting high-risk parents should adopt an ecological perspective addressing problems at multiple levels including modules targeting:

- the individual parents;
- family relationships (parent-child and marital relationships); and
- the social context of the family.

The PUP program includes modules that intervene in each of these three domains.

In developing the PUP program, it was acknowledged that the intervention should be flexible as there would be considerable variability in the needs of individual families. Therefore the format of the program was conceptualised as a structured, non-sequential (SNS) intervention. That is, the intervention is structured insofar as a module in the manual addresses each identified problem area. However, the program is non-sequential in that the order of presentation is determined by taking into account the initial assessment and the salient presenting problems for the family in each session.

It is common for high-risk parents to report a life crisis or stressor of such magnitude that their attention to a structured session is minimal, reducing the likelihood they will benefit from the session. In order to respond appropriately and flexibly in such situations, the therapist in each session determines the immediate stressor for the parent and addresses these issues by developing action plans. Through this process, the therapist helps the parent acquire proactive problem solving skills to cope more effectively with life stressors. Over the course of the program there is an increasing emphasis on parents taking responsibility for executing these action plans and developing their own action plans. In this way, the presentation of life crises during the intervention is anticipated and used as a therapeutic opportunity.

The theoretical background of the PUP program has implications for the methodology used to evaluate the program.

Evaluating the PUP program using single case methodology

Evaluating the effectiveness of the PUP program began with a series of single case studies to ascertain whether parents would engage in the program and would benefit from this involvement. Families who attended a methadone maintenance clinic were offered the opportunity to take part in a pilot program called PUP. Following a comprehensive assessment of parental factors including psychological problems and drug use, child factors such as behaviour problems, and parent-child relationship, each family took part in the 12-session PUP program.

The program was delivered in clients' homes. Of nine families who began the program, eight were recontacted at three months. All families reported significant improvements in three domains: parental functioning, parent-child relationship and parental substance use and risk behaviour. Of particular interest was the decrease in concurrent alcohol use and HIV risk taking behaviour. The families reported high levels of satisfaction with the program, mirroring the findings above. The results of this series of single case studies were sufficient to conclude that the parents were willing to engage in the program and benefited from the program.

While the methodology of the pilot study provided evidence suggestive of a positive impact, it is not possible to conclude that the changes were not due to maturation or extraneous variables. For example, the parent's motivation to attend the program may have been influenced by transient problems in the family that would have resolved spontaneously without intervention. A subsequent study employed Randomised Controlled Trial methodology to further investigate the internal reliability of the program.

Comparison conditions and randomisation

The intensive PUP program was compared to a brief two-session intervention and to standard care. Participants were randomly allocated to one of the three conditions immediately following screening for eligibility according to the CONSORT guidelines (Moher et al., 2001).

Treatment integrity

The Structured Non Sequential approach adopted by the PUP program allows for flexibility in the delivery of the PUP program, but increases the complexity of demonstrating that the program was delivered according to the manual. Rather than simply demonstrating that therapists competently implemented each module, it was necessary to ensure that the program delivered was based on an accurate assessment and formulation of the family's problems. To measure the quality of program implementation under these conditions, PUP has adapted the approach described by Henggeler and his colleagues in their evaluation of Multi-Systemic Therapy (MST) (Henggeler, Schoenwald, Liao, Letourneau & Edwards, 2002a).

MST shares characteristics in common with PUP insofar as the intervention is manualised but the intervention is tailored according to the needs of the family. Rather than follow a fixed format, MST therapists are expected to adhere to nine general treatment principles (Henggeler, Schoenwald, Rowland & Cunningham, 2002b). The quality of program administration is measured in terms of the therapist's success in adhering to these nine principles (Henggeler et al., 2002a). It was argued that supervision was important in maintaining therapist adherence and scales were developed to measure therapist and supervision adherence to the MST principles (Henggeler et al., 2002b). Together with feedback from the families engaged in treatment, these scales make up the MST quality assurance system (Henggeler et al., 2002a).

Issues of measurement

In line with the theoretical orientation of the PUP program that functioning in the parenting role is influenced by multiple domains, instruments were used to measure:

- each child's emotional, social and behavioural functioning;
- parental psychological functioning and attitudes to parenting;
- parent-child interaction; and
- the impact of social contextual factors, such as life events or availability of support, on the parent's perception of stress in the parenting role.

Details of the actual measures used are provided in the PUP treatment manual (Dawe et al., 2000).

Major findings

Families who took part in the PUP program reported substantial improvements across almost all domains. They reported decreases in parental stress, decreases in their child abuse potential generally and a reduction in rigid parenting attitudes. They also reported that their child(ren) were demonstrating fewer behavioural problems and had improved in their peer relationships. These changes were evident at the three-month post-treatment assessment point and are enduring at six-month follow up. By way of contrast neither the Brief Intervention nor the Standard Care groups showed any improvements (Dawe, Harnett, Rendalls & Staiger, 2003).

PUP dissemination study

Effectiveness or dissemination studies have different aims from efficacy studies. Specifically there is a change of focus from what is possible under ideal conditions to what actually happens in the real world.

The PUP dissemination study is currently being conducted in methadone clinics in NSW. The primary focus of the PUP dissemination study is to assess the success of the program in engaging families to participate in the program. This includes:

- evaluating a model of training and supervision;
- assessing the quality of program implementation when the intervention is delivered by practitioners in clinical settings; and
- evaluating the effect of the program on client outcomes, including client satisfaction with the program.

Training was provided to practitioners from different disciplines working in the methadone clinics. As there was variability amongst the practitioners in their prior training in parenting interventions, it was important to ensure that each practitioner received an optimal level of training and supervision to be confident and competent to deliver the PUP program.

In real-world settings it is acknowledged that time release for training is costly and can involve a great deal of organisation to facilitate attendance. Redundancy in training is costly and can reduce the uptake of programs if the initial investment of time for training is perceived to be too high. On the other hand, the program may be less effective if inadequately trained professionals fail to implement it competently (Clarke, 1995; Weisz et al., 1995). To determine that training is effective, measures are being used to evaluate both the immediate outcomes of training, including the trainees' satisfaction with training, and more distal outcomes such as the subsequent quality of program implementation and client outcomes. As in the efficacy study described above, Henggeler and colleagues' MST quality assurance model is being used to evaluate the quality of supervision and the therapists' delivery of the program (Henggeler et al., 2002a). The measures used in the RCT to measure changes in the parents are also being used in the dissemination study to allow a comparison between the changes that occurred in the PUP program under optimal conditions compared to those found in real-world settings.

Conclusions

There are many ways to evaluate intervention programs. A general understanding of the different approaches to evaluation is useful for practitioners wanting to adopt an evidence-based approach to clinical practice. Understanding the strengths and limitations of clinical research evaluations allows practitioners to critically appraise the usefulness of an intervention developed and implemented elsewhere with their own client group and presenting problems. An understanding of clinical evaluations, and the methodology of single case designs, is useful in providing the most convincing evidence for the effectiveness of their clinical work. It is hoped that the overview of research methodology presented in this chapter will assist practitioners to critically appraise the literature of treatment evaluations, and also provide information on methods for evaluating the interventions employed in their own clinical practice.

References

- Abidin, R. R. (1995). *Parenting Stress Index* (3rd Ed.). *Professional Manual*. FL: Psychological Assessment Resources.
- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th Ed.). Washington, DC: American Psychiatric Association.
- Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th Ed.). Upper Saddle River, NJ: Prentice Hall.
- Barker, C., Pistrang, N., & Elliott, R. (1994). *Research Methods in Clinical and Counselling Psychology*. Chichester, England: John Wiley.
- Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., et al. (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*, 276(8), 637-9.
- Belsky, J. (1984). The determinants of parenting: A process model. *Child Development*, 55(1), 83-96.
- Blampied, N.M. (2001). The third way: Single-case research, training, and practice in clinical psychology. *Australian Psychologist*, 36(2), 157-63.
- Chambless, D.L., & Hollon, S.D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 66(1), 7-18.

- Cicchetti, D., & Toth, S.L. (1998). Perspectives on research and practice in developmental psychopathology. In W. Damon, I. Sigel & K.A. Renninger (Eds), *Handbook of Child Psychology: Vol 4. Child Psychology in Practice* (5th Ed.) (pp. 479-583). New York: Wiley.
- Clarke, G.N. (1995). Improving the transition from basic efficacy research to effectiveness studies: Methodological issues and procedures. *Journal of Consulting and Clinical Psychology, 63*(5), 718-25.
- Dadds, M.R., & Sanders, M.R. (1992). Family interaction and child psychopathology: A comparison of two observation strategies. *Journal of Child and Family Studies, 1*(4), 371-91.
- Dawe, S., Harnett, P.H., & Rendalls, V. (2000). *The Parents Under Pressure Program: Facilitator's Manual (Substance Misuse)*. Brisbane: Griffith University.
- Dawe, S., Harnett, P.H., Rendalls, V., & Staiger, P. (2003). Improving family functioning and child outcome in methadone maintained families: The Parents Under Pressure program. *Drug and Alcohol Review, 22*, 299-307.
- Dawe, S., Harnett, P.H., Staiger, P., & Dadds, M.R. (2000). Parent training skills and methadone maintenance: clinical opportunities and challenges. *Drug and Alcohol Dependence, 60*, 1-11.
- Dudley, J.R., & Stone, G. (2001). *Fathering at Risk: Helping Nonresidential Fathers*. New York, NY: Springer Publishing Co.
- Eyberg, S.M., & Robinson, E.A. (1983). Dyadic parent-child interaction coding system (DPICS): A manual. *Psychological Documents, 13*(2), MS. 2582, 24.
- Fagan, J., & Hawkins, A.J. (Eds) (2001). *Clinical and Educational Interventions with Fathers*. Binghamton, NY: Haworth Clinical Practice Press.
- Gaynor, S.T., Baird, S.C., & Nelson Gray, R.O. (1999). Application of time-series (single-subject) designs in clinical psychology. In P.C. Kendall & J.N. Butcher (Eds), *Handbook of Research Methods in Clinical Psychology* (2nd Ed.) (pp. 297-329). New York, NY: John Wiley & Sons Inc.
- Hayes, S.C., Barlow, D.H., & Nelson Gray, R.O. (1999). *The Scientist Practitioner: Research and Accountability in the Age of Managed Care* (2nd Ed.). Needham Heights, MA: Allyn & Bacon.
- Henggeler, S.W., Schoenwald, S.K., Liao, J.G., Letourneau, E.J., & Edwards, D.L. (2002a). Transporting efficacious treatments to field settings: The link between supervisory practices and therapist fidelity in MST programs. *Journal of Clinical Child Psychology, 31*(2), 155-67.

- Henggeler, S.W., Schoenwald, S.K., Rowland, M.D., & Cunningham, P.B. (2002b). *Serious Emotional Disturbance in Children and Adolescents: Multisystemic therapy*. New York, NY: Guilford Press.
- Hsu, L.M. (1998). Random sampling, randomization, and equivalence of contrasted groups in psychotherapy outcome research. In A. Kazdin (Ed.), *Methodological Issues and Strategies in Clinical Research* (2nd Ed.). Washington, DC: American Psychological Association.
- Kazdin, A. (1981). Drawing valid inferences from case studies. *Journal of Consulting and Clinical Psychology*, 49, 183-92.
- Kendall, P.C., Butcher, J.N., & Holmbeck, G.N. (Eds) (1999). *Handbook of Research Methods in Clinical Psychology* (2nd Ed.). New York, NY: John Wiley & Sons Inc.
- Kendall, P.C., & Chu, B.C. (2000). Retrospective self-reports of therapist flexibility in a manual-based treatment for youths with anxiety disorders. *Journal of Clinical Child Psychology*, 29(2), 209-20.
- Milner, J.S. (1986). *The Child Abuse Potential Inventory* (2nd Ed.). North Carolina: Psytec Inc.
- Mitchell, M., & Jolley, J. (1996). *Research Design Explained* (3rd Ed.). Orlando, FL: Harcourt Brace College Publishers.
- Moher, D., Schulz, K.F., & Altman, D.G. *Checklist of Items to Include When Reporting a Randomized Trial* (56-8). Retrieved on 11 February 2003, from <http://www.consort-statement.org/revisedstatement.htm#checklist>
- Moher, D., Schulz, K.F., & Altman, D.G. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*, 357(9263), 1191-4.
- O'Gorman, J.G. (2001). The scientist-practitioner model and its critics. *Australian Psychologist*, 36(2), 164-9.
- Pipp Siegel, S., & Biringen, Z. (1998). Assessing the quality of relationships between parents and children: The emotional availability scales. *Volta Review*, 100(5), 237-49.
- Raimy, V.C. (1950). *Training in Clinical Psychology*. Englewood Cliffs, NJ: Prentice-Hall.
- Roberts, M.W., & Hope, D.A. (2001). Clinic observations of structured parent-child interaction designed to evaluate externalizing disorders. *Psychological Assessment*, 13(1), 46-58.
- Sanders, M.R. (1995). *Healthy Families, Healthy Nation: Strategies for promoting family mental health in Australia*. Brisbane: Australian Academic Press.

- Sanders, M.R., Turner, K.M.T., & Markie Dadds, C. (2002). The development and dissemination of the Triple P-Positive Parenting program: A multilevel, evidence-based system of parenting and family support. *Prevention Science, 3*(3), 173-89.
- Shapiro, M.M. (1967). Clinical psychology as an applied science. *British Journal of Psychiatry, 113*, 1039-42.
- Shapiro, M.M. (1985). A reassessment of clinical psychology as an applied science. *British Journal of Clinical Psychology, 24*, 1-11.
- Stricker, G., & Trierweiler, S.J. (1995). The local clinical scientist: A bridge between science and practice. *American Psychologist, 50*(12), 995-1002.
- Trierweiler, S.J., & Stricker, G. (1998). *The Scientific Practice of Professional Psychology*. New York, NY: Plenum Press.
- Turpin, G. (2001). Single case methodology and psychotherapy evaluation: from research to practice. In C. Mace & S. Moorey (Eds), *Evidence in the Psychological Therapies: A critical guide for practitioners* (pp. 91-113). New York, NY: Brunner-Routledge.
- Waltz, J., Addis, M.E., Koerner, K., & Jacobson, N.S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61*(4), 620-630.
- Weisz, J.R., Donenberg, G.R., Han, S.S., & Weiss, B. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63*(5), 688-701.



3 4067 02826 9024