Evaluating the Effectiveness of Interventions in Child Psychiatry

Peter Fonagy, PhD, D Clin Psy, FBA

Objective: To review the nature of outcome measures used to evaluate psychosocial treatment in children.

Method: Recent research findings that highlight the importance of outcome data and justify randomized controlled trials (RCTs) are reviewed.

Results: Outcome measures of child therapy can focus on changes in the child's symptoms, adaptation to the psychosocial environment, cognitive or emotional capacities, and changes in interaction with others. They can also assess service utilization, related costs, and the suffering of those associated with the child before and after treatment. RCTs remain the optimal methodology, notwithstanding current critiques.

Conclusion: Studies assessing the effectiveness of psychosocial treatment must include a definition of various outcomes and use appropriate methodologies to measure them.

(Can J Psychiatry 1997;42:584–594)

Key Words: outcome, randomized controlled trials, psychotherapy, methodology

In a recent review of the effectiveness of psychosocial treatment interventions, commissioned by the United Kingdom Department of Health, Drs A Roth, M Target, R Woods, and I examined the existing outcomes literature with the hope of guiding purchasers, managers, and clinicians in their day-to-day work. The report of this review, entitled “What Works for Whom?” was in turn peer-reviewed by nearly 30 international experts before it received endorsement by the Department of Health and was published as a book (1).

The Importance of Outcome Data

There have been major changes to the mental health care delivery systems of most Western industrialized nations over the past decade. The specific modifications have, in large measure, reflected the diversity of preexisting systems—professionally controlled private services in the United States have now been brought under regional regulation by managed care (2,3), state insurance controlled services in the United Kingdom have recently been transformed into a “pretend” (internal) market (4), and as I understand it, the Canadian single-payer system may be under threat (5). The upshot of the changes has been a move toward greater accountability of the professionals to the funders and purchasers of health care. The emphasis on outcomes is a direct consequence of this shift in power. Pressures associated with reimbursement,
managed care, and fiscal stringency in general exercise a powerful influence on clinical practice and practice guidelines (6,7). I am reminded of the statement attributed to the chief executive officer of a major American managed health care organization: “In God we trust; from everybody else we demand outcome data.”

The emphasis on the evaluation of services is society’s response to the massive increase in the cost of health provision in general and mental health service provision in particular (8). Knowing more about “what works for whom” (1) may have significant advantages for user and professional alike, but these considerations are not uppermost in the minds of those promoting evidence-based medicine (9). The recent upsurge of interest in outcomes assessment is more likely attributable to the motivation to impose social and political controls on the professional practitioner than to bring about improvement in patient care. If society so desired, it could impose significant control over trial lawyers by insisting on randomized controlled studies for a variety of legal procedures. It would be fascinating to see a cost-benefit analysis of “plea bargaining” versus “not guilty pleas” using randomized assignments.

This is not to say that insistence on treatment regimens “of known effectiveness” is a necessarily adverse outcome of social and structural changes. There were major shortcomings in systems of mental health care delivery that can be clearly attributed to the absence of adequate interest in research findings. It is of more than passing concern that most pediatric psychopharmacological treatments have not been thoroughly evaluated and are indeed not all approved by the Food and Drug Administration for use with children (10). There are evidently many treatments in common use that are in all probability not efficacious. It is an illusion that clinical experience can tell us what is efficacious. Any historian of medicine would delight in enumerating dramatic counter-examples, such as insulin coma therapy, but this is not just an issue of yesteryear. Ethically and professionally, it seems essential to eradicate treatments that are likely to be wasteful of scarce resources.

Most would agree that, notwithstanding genuine scientific and clinical considerations, the interactions between social anxieties and professional concerns, that is, public demand and clinical opinion, will determine if certain interventions will continue to be offered. For example, interventions for substance misuse and self-harm are not particularly effective (11,12), yet their elimination would create an intolerable clinical gap and could undermine the possibility of the discovery of new interventions and effective management strategies. We shall return later to the key issue of the dialectic between petrification consequent upon rigid outcome controls and a free-for-all where such outcome information is ignored.

What Is “Outcome” or “Whose Outcome Is It Anyway”?

It should be self-evident, but it hardly is, that child mental health outcomes cannot be considered in absolute terms. First, even in the arena of high technology medicine, it is now generally agreed that the impact of any intervention must be seen in the context of the totality of that person’s life experience, ostensibly quantified as quality-adjusted life years (QALYs). Demographic changes, in particular the rising proportion of older individuals in society, have led to profound questions about the circumstances under which death may be considered a positive rather than a negative outcome. When is the preservation of a family unit a positive outcome, or rather, for whom is that outcome positive—the child, the parents, the clinician, or the purchaser who would be required to fund alternative care? What if the outcomes diverge: what is good for the child is less favourable to the family or to the service provider? We should be aware of the old adage, “In most instances the most cost-effective intervention is to do nothing.”

Psychotherapy researchers are particularly conscious of the danger of imposing ethnically rooted cultural biases on what is designated as “needing treatment” and “good outcome” (13). For instance, the achievement of selfhood through the separation-individuation process is one of the cornerstones of psychotherapeutic interventions. Yet is Lasch (14) correct that the emphasis on individual achievement in Western culture is excessive and that an appropriate submission to the goals of the family and community (15) may be a far better indicator of healthy adaptation? Such differences are particularly acute in the area of child development and parenting. Rogler (16) outlined some of the practical steps that culturally sensitive outcome research requires. In particular, it is important to ensure that interventions are consonant with the subjective culture of the ethnic group to which they are applied and that instruments used are able to integrate cultural meanings with the pertinent scientific categories.

A further consideration that complicates the issue of outcomes for childhood interventions is the developmental framework, which has appropriately come to dominate the fields of child psychiatry and psychology (17). A clear implication of developmental psychopathology is that symptoms cannot be considered the sole or even the most important criteria of treatment effectiveness. If psychiatric disorder is not only the end result of a complex series of interactions across time of biological, social, and psychological characteristics but is itself part of a complex transactional causal chain, good outcome might sometimes be an increase rather than a decrease in symptomatology.

In our work on child therapy outcome, we have conceptualized 5 levels of outcome: First is the symptomatic or diagnostic level. It is at this level that measurement is most sophisticated. Both categorical (18–22) and dimensional
assessments may be achieved reliably with relative ease, and the computerization of such assessments will overcome many remaining practical problems. The concern at this level is the low degree of observed agreement between informants, especially concerning internalizing behaviours (25–27). Achenbach (28) recently offered a decision-tree approach for making taxonomic assignments from multiple sources of data. This approach, if applied, would enable the researcher to distinguish 3 reasons for such disagreements: 1) the contextual dependence of the child’s problem; 2) comorbidity or multiple syndromes; and 3) distorted perception on the part of some informants. The value of this diagnostic strategy in outcome research remains to be explored.

A broad range of symptomatic assessments is essential since about half of individuals in community samples who meet diagnostic criteria for a DSM-IV disorder are likely to meet criteria for another disorder as well (29,30). In clinical samples, the prevalence of comorbidity is likely to be even higher (70% to 75%) (31). Problems tend to go together. For example, delinquency is commonly associated with drug abuse (32–34). The link exists either because they serve similar functions in relation to development and adaptation (drug abuse and delinquency may both be understood in terms of attachment-related considerations [35]) or because one represents a risk factor for the other (for example, reading retardation may be a risk factor for conduct problems [36]). Without evaluating the full breadth of an individual’s symptoms, we will have limited information on the effectiveness of a treatment except for the individual’s primary diagnosis.

The second level of measurement concerns adaptation to the psychosocial environment. Mental health problems impinge on many domains of the child’s functioning, yet many treatments are evaluated solely in terms of their impact on core symptoms. Recent metaanalyses demonstrated that such limited evaluations lead to a potentially misleading overestimation of the effect of the treatment, at least in terms of effect size (37). We know that many treatments leave important areas untouched. For example, stimulants do not benefit the academic performance and peer relations of children with attention-deficit hyperactivity disorder (ADHD) (38).

In this respect, our available measurement options are limited. Target and Fonagy (39) identified 14 dimensions that may be reliably assessed on the basis of a parental interview. This method proved to be superior to simple global measures of functioning such as the Child Global Adjustment Scale. Unevenness of adaptation may be characteristic of our clinical populations (40). The presence of prosocial attributes may be particularly important in the comprehensive assessment of outcome. There exist other operationalizations of this dimension, for example, Hodges’ Child and Adolescent Functioning Assessment Scale interview (41). A strength of our approach is the developmental anchoring of adaptation. We have succeeded in operationalizing anchor points for specific age ranges (2 to 3 years, 4 to 5 years, 6 to 9 years, 10 to 13 years, and 14 to 18 years) and can thus describe a child in terms of a clinician-rated developmental profile.

The third level concerns mechanisms, the cognitive and emotional capacities that probably underpin both symptomatology and adaptation. Cohen (42) provided a cogent case for why outcome research must be grounded in a theory of both normal and abnormal child development as well as a theory of therapeutic action. He implied that outcome measures should be pertinent to the developmental theory within which a treatment is rooted and should satisfy the practical concerns of patients, families, clinicians, and funders. A good example of the value of this approach is a recent randomized, double-blind, placebo-controlled trial of methylphenidate (43). These investigators explored 2 capacities ostensibly underlying ADHD: response inhibition and response reengagement. They administered 3 doses of the drug and found different dose–response functions for these 2 capacities. Whereas response reengagement had a linear relationship to dose, response inhibition was optimal in lower or middle doses. This finding argues against a typical clinical practice of determining the response to a stimulant from a single measure, such as a parent report of child behaviour.

Capacities important for the developmental outcome of psychological disorder include the assessments of dimensions such as affect regulation (44), understanding of emotions (45), self-representation (46), understanding of mental states in self and others (47), forming emotional bonds (48), making moral judgements (49), and attributional biases (50,51).

It should be noted that while both symptoms and adaptation may be directly assessed from informants’ reports or observation of behaviour, capacities such as understanding of affect must be inferred from the child’s performance on specially designed tasks. This creates measurement problems and challenges; for example, researchers are required to distinguish between competence and actual performance. What the child is capable of and what he or she tends to do are often rather different. The theoretical framework of the clinician may determine the dimensions identified as critical to specific interventions. An adequate demonstration of effectiveness must include measuring change in those capacities (or mental processes) which the clinician believes may underpin or place the child at risk of symptomatology or maladaptation. Surprisingly, outcome studies measuring change at this level frequently fail to demonstrate that symptomatic change was associated with modifications of underlying mechanisms. For example, cognitive–behavioural therapy is sometimes successful without demonstrable changes in cognitive structures (52).

The fourth level is transactional. Child mental health professionals and developmentalists have traditionally been concerned with the contextual influences on the child’s adaptation and generally see progression and mutual accommodation between the individual and changing environments (53).
Since developmental psychopathology posits transactional interactions between the mental state and behavioural predispositions of the child and the reactions of the environment to it across time (17), it seems to us essential that evaluators should assess the quality of these transactions. Research on conduct disorder over the past 10 years has indicated how risk factors such as temperament (54,55) and parents’ personal and interpersonal problems, for example, maternal depression (56), may interact to cause elevated noncompliance (57). The caregiver’s failure to cope with the oppositional behaviour of the child may be further aggravated by the absence of social support (58) and the high level of psychosocial stress associated with the environment in which the family lives (59). Thus the contextual influences that may have transactional relations with the child’s problem include the parents, family relations, characteristics of the community, and the child’s school, as well as more general cultural factors (60). Effective treatments address these risk factors, which must be evaluated alongside the child’s characteristics.

A good example of the pertinence of the transactional dimension for outcome studies comes from family-based treatments of schizophrenia. High levels of expressed emotion (parents’ hostility, criticism, and overinvolvement) predict adolescent-onset schizophrenia (61) as well as relapse after discharge (62). A number of well-conducted outcome studies have demonstrated that intervention programs addressing expressed emotion reduce the likelihood of relapse in individuals on medication (63–66).

Many effective interventions are directly aimed at bringing about change at the level of transactional processes. For example, parent training and social skills development programs for conduct problems of the kind described by Patterson (67) and others (68,69) are effective in improving child management skills (70), increase prosocial behaviours (71), and reduce behavioural problems in middle childhood (72).

It is clearly also possible for symptomatic improvements in the child to bring about negative as well as positive changes in the family, school, or the wider social environment. For example, Szapocznik and others (73) compared structural family therapy with psychodynamic therapy for disruptive behavioural disorders. On parent-rated measures of child behaviour, both treatments led to change that was maintained at one-year follow-up. The assessment of family functioning (transactional level), however, indicated that whereas family functioning had deteriorated in the individual group by the one-year assessment, it continued to improve in the family therapy group.

The final level of outcome concerns the level of service utilization. A number of interventions make strong assumptions concerning a reduction in posttreatment service utilization. For example, home-based preventive interventions implemented in early childhood may have the power to reduce childhood maltreatment and thus lessen the pressure on child welfare services. A number of studies offer evidence that maltreatment can be prevented (74–78), although not all studies demonstrate significant benefit (79,80). MacMillan and others (81,82) have reviewed this literature and concluded that “extended home visitation can prevent child physical abuse and neglect among families with one or more of the risk-markers of single parenthood, teenage parents status and poverty” (81, p 854).

A methodologically somewhat weaker study of multimodal treatment for ADHD (83) demonstrated a fourfold reduction in felony arrests 9 years posttreatment in a group of children offered such treatment in addition to stimulant medication when contrasted with a comparison group who received medication alone.

Service level outcome may also be conceptualized at the level of general service provision and the quality of integration of various services. Thus an important outcome of intensive case management (84,85) may be the better integration and greater accessibility of services relevant to child mental health. The development of standardized measures of service utilization has hastened the development of assessment tools in other areas. The reason for this is most likely the large variability of current service provision and the absence of adequate communication between service providers for children and their families.

This multiple-level outcome system has many features in common with the model advanced recently by Hoagwood and colleagues (Symptoms, Functioning, Consumer Perspectives, Environments and Systems) (86). These workers, however, consider consumer satisfaction as a separate dimension of outcome, although they combine adaptational and cognitive and emotional capacities into the single domain of functioning. Both systems highlight that outcomes should be considered across a range of domains and underscore ways in which the measurement of child mental health outcomes is both more complex and yet less well developed than the assessment of adult functioning. Sophisticated evaluations of outcome are essential if we are to address the issue of the appropriateness of a particular treatment for a particular individual. It is likely that particular profiles of scores on the 5 dimensions specified at baseline might predict the differential efficacy of the treatment for patient groups.

Justifying RCTs

One issue that has to some degree polarized the field has been the controversy concerning the need for RCTs. Researchers in the field hold up the RCT as “a gold standard” of evaluation research and argue that observational methods provide no useful means of assessing the value of therapy (87). Others argue cogently against this (88,89). The logic of RCTs is unassailable; their superiority to observational methods is so self-evident that alternative strategies can only be justified in terms of the limitations of RCTs. The limitations of RCTs can be seen as due either to the inherent nature of
the method—this would be a limitation in principle—or to the conduct of the experiments—a limitation of procedure (90). Most of the objections to RCTs tend to concern limitations in the procedure, which, at least in theory, could be overcome. We shall return to these when we consider the outcome field from the point of view of psychosocial and physical treatments, since such issues should be addressed in relation to specific studies rather than the experimental approach as a whole. The question to ask at this point about issues of principle is: Should RCTs be replaced by nonexperimental observational methods because in the child field they are unnecessary, inappropriate, impractical, or inadequate? Black, in a recent article in the British Medical Journal (90), has helpfully summarized 4 main reasons why observational methods may be needed.

1) **RCTs may be unnecessary** when the effect of an intervention is so dramatic that the likelihood of unknown confounding factors being important is appropriately considered negligible. Observational studies are adequate to demonstrate the effectiveness of insulin for diabetics, the immobilization of fractured bones, and anesthesia for operations. If only the effects of child psychiatric interventions were this obvious. Few would argue that any child mental health intervention can be appropriately considered in this category. The major metaanalytic studies in the field have estimated the effect size of psychological therapies to be 0.7 to 0.8 (91–94), with the most recent analysis yielding a moderate effect size of 0.54 (37). Thus only about 60% of patients treated are likely to be better off than untreated controls. The effects are too small, therefore, to suggest that RCTs of child mental disorder treatments are unnecessary.

2) **RCTs may be thought to be inappropriate** under 3 conditions:

(a) The phenomenon under observation is so rare that it is unlikely to occur in the context of a controlled trial. This may be the case with interventions that are designed to prevent a relatively rare event, such as crib death. This point is relevant to prevention intervention RCTs that attempt to address such rare disorders as childhood autism (95). Sadly, prevalence studies of child psychiatric disorders have produced relatively high prevalence estimates in community samples for most disorders (96–98). Thus, in most cases, RCTs cannot be ruled out on the basis of rare occurrence.

(b) Experimentation may be inappropriate in principle for trials when the outcomes of interest are far in the future. The question may receive no answer during the time period within which it seems of importance. So-called sleeper effects from child psychotherapy trials (99,100) alert us to the possibility that the impact of therapy may not be present at the end of treatment, particularly for interpersonal or psychodynamic therapies. This suggests that extended follow-up periods should be the norm (which presents a procedural problem), but there is also an issue of principle here. The more extended the follow-up is, the more difficult it becomes to ascribe any changes observed to the original treatment; patients will often seek further treatment in the intervening period (101), and the relative impact of the treatment in the context of the child’s life experiences may be expected to decrease over time. Thus a very long-term follow-up may be hard to interpret. It is doubtful, however, whether observational studies could clarify this picture. Intercurrent treatments create the same problems for observational studies as they do for RCTs. Thus the issue here is not whether RCTs are made inappropriate by sleeper effects but rather if changes uniquely attributable to that treatment are ever genuinely knowable. Surely the answer lies in developmentally sophisticated studies in which the target for outcome includes risk and protective factors, as well as primary symptoms.

(c) Brewin and Bradley (102) argue that random allocation may cut across the sometimes undeclared preferences of both clinician and patient, and RCTs are inappropriate when such considerations could enter into treatment administration (for example, psychotherapy). Such a built-in resistance to a new treatment may show the intervention to be less efficacious than it actually is. Globally, however, the evidence points the other way. Although the studies are few in number, the dozen or so that are available indicate that the effectiveness of psychotherapy in a research setting is significantly larger than in a clinic setting (93,103). It is not yet known if patient mix, the breakdown of the intervention, or inadequate assessment of outcome in the clinic setting is primarily responsible for this. There is some indication from metaanalytic studies that effect sizes are significantly greater when the researcher’s orientation coincides with the therapy investigated in the RCT (37,91). This again is an issue of procedure, rather than principle, as a number of studies have demonstrated that crossover of therapists across psychosocial treatments (for example, the National Institute of Mental Health Multisite Affective Disorder Study) is practicable. RCTs are appropriate to child mental health to the extent that therapists may be trained in all the treatment interventions which are part of the random allocation. In any case, the direction of evidence points away from this line of argument.

3) There are a number of serious practical obstacles that all RCTs have to face, and at least some of these may be impossible to confront in child mental health RCTs. If RCTs are impossible, clearly an issue of principle is being raised.

As an example, clinicians may refuse to participate because of a deep personal conviction that makes them reluctant to alter their practice even though the profession as a whole accepts that the appropriateness of a specific intervention is questionable (104). Thus there may be “group equipoise” about psychodynamic psychotherapy, but individual psychodynamic therapists may feel that any other treatment is unethical. In my experience, this situation is most likely to occur when clinicians anticipate outcomes from a randomization that will be unfavourable to their current practice. The researcher has to be alert that these forms of resistance should
not subvert the RCT (105). Given the wide range of psychosocial treatments currently available in child psychiatry (106), the absence of “individual equipoise” should be confronted rather than accepted as inherent to the field.

The second issue that frequently emerges as if it was one of principle is the issue of ethics. It is certainly unacceptable to withhold a treatment of known effectiveness from a child. Metaanalytic studies have already demonstrated that, at least in the research context, a child receiving psychotherapy is likely to be better off than the vast majority of untreated controls (37,91–94). Some have argued on the basis of this that untreated control groups should never be used, even in adult psychotherapy research (107). The problem is more acute in the pediatric context. Children may be going through critical developmental phases that could be compromised if their disorder is allowed to continue untreated. Even if a disorder is likely to remit spontaneously over a few years, as is the case for school phobia (108), it may, within that time, seriously disrupt the child’s educational and social experience; family functioning may be distorted if a parent is kept at home by the child’s symptom.

To counterbalance such an argument, we should be aware that only a relatively small proportion (10% to 20%) of children identified as having significant psychiatric disorder find their way to mental health professionals (25,109–111); thus treatment as usual is no treatment at all. It is therefore hard to prove that untreated children are inevitably greatly disadvantaged relative to the vast majority of individuals with comparable levels of the disorder. Further, controlled studies often inadvertently demonstrate that so-called placebo treatments are as effective and sometimes more effective than their active-treatment counterparts (112,113). The continued practice of therapies of unknown effectiveness seems to me as ethically questionable as the temporary withholding of these interventions. The use of minimal-treatment (for example, parent guidance) control groups may provide a practical solution to many of these ethical dilemmas.

A more realistic practical objection to experimentation is the sheer scale of the task. With the growth of client-centred and other existential therapies in the 1940s and 1950s, behaviour therapies in the 1960s, and systemic family work and cognitive–behaviour therapy in the 1970s and 1980s, there has been an explosion of so-called models of therapy to the point where the number of different forms of therapy currently offered probably exceeds 200 for children alone (114). Not only are types of psychotherapeutic intervention large in number, but most of these have many components. For example, cognitive therapy tends to have behavioural as well as purely cognitive components. It seems that it will never be practical to subject but a very limited subset of available treatments to systematic evaluation (115).

The counterargument here is that the diversification of therapies is illusory. Since the 1960s, there has been a strong integrative movement in the adult psychotherapy field (116–118). The proportion of psychotherapists who consider themselves as eclectic has been estimated between 30% and 68% (119). Eclecticism, however, is poorly justified by research, only casually integrated with theory, and may in many instances be an indication of inadequate training or lack of thorough knowledge (120).

A complementary approach is one of integration, the identification of common factors across psychotherapies (121). There have been many thoughtful attempts at integration (122). The lack of clear differences among different modes of treatment is often attributed to the absence of genuine treatment mode-specific factors. Because psychosocial therapies share numerous situational constraints (for example, therapeutic partnership, hope on the part of the client, and strengthening of self-efficacy), it should not surprise us that important common variables are highly likely to emerge. These so-called nonspecific factors naturally raise questions concerning the nature of the therapeutic process and the ways in which aspects shared by treatments could be the key to understanding change in therapy (123). Thus the actual number of genuinely different approaches may be far smaller than the number of claims to uniqueness.

A strong case could be made that the results of experimental outcome studies are of limited importance unless or until the question of integration has been properly addressed. Over 50% of clinicians offer therapies that borrow elements from different orientations on a “mix and make” basis; it should not surprise us, then, that the average clinician has limited interest in the evaluation of therapies in their pure form. Yet the permutation of components of even 2 therapies is so large that an empirical approach to this question is unlikely to be arrived at in the near future. Eventually, theoretical orientations will have to be integrated since they are all approximate models of the same phenomenon: the human mind in distress. There are indications that cognitive therapists are becoming increasingly interested in nonconscious mental processes (124–129) and relationship aspects of their treatment (130,131). Conversely, psychodynamic therapists are more concerned about the nature of knowledge representation and deficits in cognitive functions, which could account for slow progress in psychodynamic psychotherapy (132–135). It is to be hoped that these are not isolated initiatives and that the “schoolism” of psychosocial interventions is now in the past. It is unlikely that there are genuinely over 200 different ways of helping children: the therapeutically effective components are likely to be far fewer in number. The identification of those components and the combinations in which they may be helpful for specific disorders should be the goal of RCTs—a goal which is eminently reachable with a concentration of existing resources.

The task of the evaluator remains sizeable but not, in principle, impossible to perform. The number of different therapies is likely to be reduced by the mere emphasis on evaluation on the part of funders. This may well have the
unintended consequence of shedding certain types of interventions from the therapeutic portfolio as clinicians decide, on pragmatic grounds, to abandon therapies for which no funding will be forthcoming. Naturally, from a scientific standpoint, this would be a regrettable outcome.

4) RCTs are often regarded as inadequate because their external validity or generalizability is low (136). This is a hotly debated issue in the field of psychotherapy research (137). There are 3 major reasons why randomized trials in many areas of health care may have low external validity.

First, participating health care professionals may be unrepresentative in terms of their enthusiasm for the treatment, the setting in which they work, or their experience. Meta-analytic studies have demonstrated the unrepresentativeness of professional participants in psychotherapy clinical trials, for example, the underrepresentation of qualified professionals (94,138). The problem here is that we cannot readily identify the therapist characteristics that affect the outcome of psychotherapy. Therapist experience, for example, has only a small (0.2 to 0.3) effect on outcome in adult psychotherapy (139). Even training appears to have only a weak relationship (140–142) to positive outcome. The pattern for child and adolescent treatments is even more puzzling. For example, Weisz and colleagues (37) found that paraprofessionals (for example, parents, teachers) treating children, but not adolescents, produced a larger mean effect size with behavioural than with nonbehavioural therapies. Further, behavioural methods of treatment for adolescents yielded greater effect sizes for student therapists than paraprofessionals. Thus, while therapists in RCTs may be unrepresentative with respect to some characteristics critical to outcome, the field is considerably far away from identifying what these vital attributes might be.

To some degree, the operationalization of therapeutic procedures in manuals of psychotherapy mitigated this threat to external validity. Individual differences between therapists are reduced by manualization of treatments (141). It will always be true that professionals, including surgeons, differ in terms of their level of competence. Appropriate sampling is a practical solution to the problem. Bias from unrepresentativeness of professionals in RCTs is, therefore, a limitation of procedure rather than principle.

A second threat to generalizability from RCTs is the unrepresentativeness of the participants. This is a major problem for clinical trials in medicine. Inclusion criteria are frequently so restrictive that patients eligible for inclusion represent but a small fraction of those to whom the treatment might be offered in normal practice. For example, one study indicated that only 4% of patients currently benefitting from heart bypass surgery in the United States would have been eligible for inclusion in the trials conducted in the 1970s (143).

Research protocols for psychotherapeutic treatment have been frequently criticized for not being sufficiently responsive to the complexity of patients’ presentations (144–146). The strongest argument concerns the reduction of generalizability brought about by the greater severity and comorbidity of clinic rather than research cases (147,148). The scope and nature of the dysfunction in clinical samples may attenuate the impact of the treatment because of family characteristics, noncompliance, chronicity, and other negative prognostic indicators. Thus it could be argued that research findings cannot be turned into clinical recommendations because the research data pertain to a less challenging population of cases. Such an argument, however, presupposes an interaction between severity and the efficacy of specific modes of therapy. This interaction should be demonstrated before this argument is accepted. While there is good evidence for a “main effect” of setting confirming the greater difficulties presented by clinic-based populations, we know of no demonstrations of a significant interaction between type of therapy and setting that would substantiate the case for not being able to generate RCTs as a matter or principle. Nevertheless, it seems advisable for RCTs to make use of clinic-based samples as far as possible.

It may also be argued that the extensive use of DSM-IV (149) as the framework for research-based service recommendations provides but an illusion of clarity, since diagnoses are not strongly related to outcome. For the purposes of the evaluation of particular therapeutic procedures, patients should be grouped not according to descriptive criteria, but by psychological functioning, which predicts responsiveness to treatment (150). It would be hard to identify a set of client attributes that apply equally to specific treatments. Psychological mindedness may be a precondition for psychodynamic psychotherapy (151), but other forms of therapy would specify quite different preconditions. Considerations of suitability for any one theoretical approach cannot be considered to be of the same order as a reliable system of classification for disorders, whatever the limitations of that classification with regard to specific orientations (152). In fact, where such studies have been performed, diagnostic considerations were found to be important predictors of suitability for psychodynamic treatment (153–155). The DSM nosology has had a vast effect on mental health research worldwide, including psychotherapy research (156). At the moment, it is hard to envision a better system than DSM that might meet the objective of providing a scientific context for recommendations regarding psychotherapy practice and training.

The third and final concern about generalizing RCTs is that the treatment offered may be atypical. In other fields of health care, there is evidence that patients who participate in trials may receive superior care, regardless of which arm of the trial they are in (157). For example, manualized psychotherapy, while necessary for research, may distort the psychotherapeutic process to a point where findings from outcome
studies are no longer relevant in the clinical setting (89). While it is undoubtedly true that experienced therapists do not use treatment manuals, it is not at all clear that this is necessarily a good thing. When research therapy programs are transferred to a clinic setting, there is likely to be an overall increase in therapeutic gain (103). There are several studies that demonstrate that greater adherence to psychotherapy protocol predicts better client outcome (140,158,159), although it should be noted that not all studies confirm this relationship (160,161). The improvement in efficacy gained in the transition between the laboratory and the clinic is an issue of current concern to many researchers (162). There is strong indication that, at least for some therapies, the adoption of laboratory-based procedures in a clinical setting may be of great advantage to the patient (163). Some have argued against treatment manuals by asserting that ordinary clinical practitioners do not see things the same as those working according to a manual, which is a quite unhelpful viewpoint which may point to a need for further professional education.

In summary, the above-mentioned studies indicate that RCTs in child psychiatry may be difficult to implement. Because RCTs are often necessary for evaluating the efficacy of specific treatments, investigators must be encouraged to use methodologies that allow for their use in clinically relevant contexts.

Summary and Conclusions

In the present paper, I have outlined some parameters that must be considered in studies which attempt to evaluate the effectiveness of interventions in children and their families. They include a clear definition of various possible levels of outcome and the choice of appropriate methodologies to measure them.

In future publications (164,165), we will examine the actual evidence that specific forms of treatment can modify psychiatric disorders commonly seen in children and adolescents and look at ways this may help us in conceptualizing future preventive public health strategies.

Acknowledgements

This manuscript was prepared as part of the National Health Service Executive (NHSE) Review of Child Mental Health Intervention Outcomes, which is the collaborative work of Dr Mary Target, Professor David Cottrell, Dr Zarina Kurtz, Dr Janet Philips, Ms Arabella Kurtz, Mr Michael Ruile, and the author. The synthesis processed here is preliminary, and the collaborative group is not responsible for the views advanced in this paper.

References


Clinical Implications

- Outcome of child psychiatric interventions should be assessed at multiple levels.
- RCTs have limitations of procedure, rather than principle, in child psychiatry.
- Most of these limitations of procedure could be addressed and accommodated in careful reviews.

Limitations

- This is a preliminary, selective report of a comprehensive review.
- The approach to conceptualizing outcome is not validated.
- Good RCTs are few in number in some important areas.
110. Tuma JM. Mental health service for children: the state of the art. Am Psychol 
108. Graham PJ. Treatment of child psychiatric disorders: types and evidence for 
107. Basham RB. Scientific and practical advantages of comparative design in psycho-

85. Koroloff NM, Eliott DJ, Koren PE, Friesen BJ. Connecting low-income families 
84. Clark HB, Prange ME, Lee B, Boyd LA, McDonald BA, Steward ES. Improving 
80. Black N. Why we need observational studies to evaluate the effectiveness of health 
77. Haggerty RJ. Reducing risks for mental disorders: frontiers for preventive 
59. August 1997 Interventions in Child Psychiatry 593
Résumé

Objectif : Examiner la nature des mesures des résultats qui servent à évaluer le traitement psychosocial des enfants.

Méthode : On examine des découvertes récentes faisant ressortir l’importance des données sur les résultats et justifiant des essais contrôlés et randomisés.


Conclusion : Les études visant à évaluer l’efficacité du traitement psychosocial doivent comporter une définition des divers résultats et faire appel à des méthodologies appropriées afin de mesurer les résultats en question.