

## Measurement and psychotherapy

### Evidence-based practice and practice-based evidence<sup>†</sup>

FRANK R. MARGISON, MICHAEL BARKHAM, CHRIS EVANS,  
GRAEME McGRATH, JOHN MELLOR CLARK, KERRY AUDIN  
and JANICE CONNELL

**Background** Measurement is the foundation of evidence-based practice. Advances in measurement procedures should extend to psychotherapy practice.

**Aims** To review the developments in measurement relevant to psychotherapy.

**Method** Domains reviewed are: (a) interventions; (b) case formulation; (c) treatment integrity; (d) performance (including adherence, competence and skilfulness); (e) treatment definitions; (f) therapeutic alliance; and (g) routine outcome measurement.

**Results** Modern methods of measurement can support 'evidence-based practice' for psychological treatments. They also support 'practice-based evidence', a complementary paradigm to improve clinical effectiveness in routine practice via the infrastructure of Practice Research Networks (PRNs).

**Conclusions** Advances in measurement derived from psychotherapy research support a model of professional self-management (practice-based evidence) which is widely applicable in psychiatry and medicine.

**Declaration of interest** The Mental Health Foundation, the Counselling in Primary Care Trust, and the Society for Psychotherapy Research (UK) (Northern Group) have supported the work of the CORE System Group.

This paper is in two parts. First, we describe recent advances in psychotherapy measurement and review the multiple perspectives currently available. Second, we argue for an increasing role for clinically responsive measurement methods that can be applied in professional self-management. Critics and supporters alike have argued in the past that psychotherapy is 'unmeasurable'. The emphasis on meaning may have been at the expense of developing methods of measurement and testable theories. Given the central place of 'measurement' in supporting evidence-based practice, it is crucial to ensure that measurement procedures in psychotherapy are well understood.

#### CONTEXT: EVIDENCE-BASED PRACTICE AND PRACTICE-BASED EVIDENCE

In support of evidence-based practice, a methodology is emerging that enables practitioners to generate clinically meaningful questions (Sackett *et al*, 1997). Many areas of medicine are acknowledged to lack evidence from methodologically robust studies such as randomised controlled trials (RCTs), and this extends to many aspects of psychiatry, including psychotherapy. The RCT paradigm has, in any case, well-recognised limitations (Kazdin, 1994). Among the technical limitations are differential attrition, non-comparability of comparison groups, psychometric problems with outcome measures, inconsistency of treatment delivered, and contamination by other treatments in trials of long-term therapy. Of greatest importance to clinicians is the poor success of RCTs in predicting outcome at the level of the individual case from data summarised at the level of group means. This has resulted in the situation where most treatments, when studied, are shown to be better than no treatment. However, little systematic advantage is

found for any particular treatment when compared with others (particularly if researcher allegiance is partialled out as a potential source of bias), and the best predictor of outcome is initial therapeutic alliance.

Despite these caveats, the evidence base for psychotherapy has been extensively reviewed. One major review (Roth & Fonagy, 1996) focuses almost exclusively on evidence from RCTs. This reveals the lack of evidence concerning many therapies, rather than evidence for or against. However, there is a complementary paradigm which is already well developed in psychotherapy (e.g. Margison *et al*, 1998). This involves gathering good-quality data from routine practice, a procedure we term 'practice-based evidence', which is discussed in greater detail below. For psychotherapy to have a robust evidence base both paradigms are needed.

#### The myth that psychotherapy is not measurable

It has been argued that there has been "a decline in theory-guided and a rise in pragmatic, clinically-oriented research" (Omer & Dar, 1992). In fact, there have been important developments in theory. These include, among others, attachment theory (e.g. Mace & Margison, 1997), sign-mediated language theory (Stiles *et al*, 1988) and theories of mind derived from developmental psychology (Hobson, 1993), the assimilation model of change in psychotherapy (Stiles *et al*, 1988) and new conceptualisations of personality (e.g. O'Brien & DeLongis, 1996).

These developments in underlying theory are crucial. However, policy also urges the improvement of measurement, as part of increased accountability within medicine and other clinical practice.

#### Stakeholders

We have considered the various 'stakeholders' along with the most salient practice-research questions in Table 1. This framework sets the context for a discussion about the most appropriate measurement strategies. Clearly, different stakeholders will have different priorities and an overall strategy is needed to coordinate these different needs (Department of Health, 1996).

<sup>†</sup>See editorial, pp. 93–94, this issue.

## Efficacy and effectiveness

The distinction between *efficacy* and *effectiveness* (Cochrane, 1972) becomes relevant when considering the different priorities of the different stakeholders. *Efficacy* refers to evidence derived from carefully designed trials where threat to internal validity is minimised. Traditionally, in evidence-based medicine the randomised controlled trial is considered as the 'gold standard' because of its ability to deal with bias through the randomisation process. However, even with the RCT there is a trade-off between rigour and generalisability (Shapiro *et al*, 1995). In particular, the more selective the sample and the more rigorously defined the intervention, the less applicable the

treatment is likely to be to routine practice. In *effectiveness* research, the design is weighted towards high generalisability, but the price is paid in greater threats to internal validity.

To resolve this dilemma, Barkham & Mellor-Clark (2000) suggest a strategy with three phases: (a) theory and generation of treatment; (b) validation and testing efficacy; and (c) dissemination and measurement of effectiveness in practice. The strategy is logical, but in practice there are difficulties to this orderly approach. First, the need for replication of efficacy studies in phase (b) is understated. Second, the time scale from inception to full implementation of a new treatment is very long, and by that time keen clinicians are likely to be already modifying the treatment. Third, the transfer

of knowledge from research studies to normal clinical practice is poor.

## MEASUREMENT STRATEGIES IN PSYCHOTHERAPY

For each of seven domains ((a)–(g) below) we give two examples of how the measures might be applied in clinical practice.

### (a) Interventions at the level of component skills

This represents the 'micro' level of analysis of a session at the smallest unit of measurement. Specific interventions are often referred to in the literature as 'verbal response modes' (VRMs), and these can be rated reliably (Elliott *et al*, 1987). The

**Table 1** Stakeholder view of evidence and psychotherapy

Stakeholder	Priorities	Salient practice-research issues	Measurement needed
Individual professional	Best use of time and resources Developing clinical skills	How to allocate time as a therapist How to improve 'skills portfolio'	Case-mix and case-load review Specific skills Adherence Competence Overall skilfulness
Patient–therapist dyad	Allocation to best treatment Monitoring progress Early identification of problems	What interventions? Why used? How accurate is the formulation? How to maintain therapeutic alliance	Case formulation Measure (personally relevant) change Case-tracking method
Peer review group	Maintaining quality within local service	Benchmarking outcomes Case complexity Treatment allocation decisions	Routine evaluation of assessment and outcome
Organisation	Managing clinical and financial risk	Treatment protocols and guidelines Identifying and managing clinical risk	Identifying outliers Risk recognition from case-tracking
Professional body	Maintaining optimal professional standards	Remediating deficits Maintaining skills Adding skills in new therapies	Specific skills Adherence Competence Overall skilfulness Ability to train others
National government research and clinical priorities	Allocation of resources between therapies Focusing on effective therapies for those in most need	Evidence-based clinical guidelines Efficacy Cost effectiveness Orderly treatment developments	Treatment definitions and manuals Quality, adherence and competence measures for each treatment Generic measures of competence and outcome Health economic measures Population-based health needs assessment methods

classifications used can be generic (not specific to the mode of therapy under study) or specific to a particular mode of therapy. The latter tend to be more reliable but operate over a narrower range of therapists' behaviour types (Elliott *et al*, 1987).

#### Clinical implications

- (i) Analysing psychotherapy at the level of the single intervention does not easily generalise to the more complex clinical skills such as formulation.
- (ii) However, there is evidence that early training at this 'skills-based' level can be effective in reducing undesirable types of therapists' behaviour, such as the excessive use of closed questions.

#### (b) Case formulation

At the 'macro' level of case analysis, formulation represents conceptualisation at the level of a whole treatment. Case formulation was initially developed in relation to psychodynamic approaches (Luborsky, 1984) and shown to be a replicable procedure. Recent work has included explicit formulation techniques in schema-focused cognitive therapy (Persons, 1989).

#### Clinical implications

- (i) Having a formulation shared with the patient can help maintain the therapeutic alliance during difficult re-enactments; or, in supervision, help understand potential re-enactments.
- (ii) The formulation can be used by the therapist after each session to ensure that the agreed focus was being maintained in that session.

#### (c) Treatment integrity

Integrity is defined simply as the extent to which treatment procedures are carried out as intended (Kazdin, 1994). The concept was intended to cover three different aspects: adherence, competence and differentiation. *Adherence* refers to the extent that the therapist is using procedures described as characteristic of the model being used. *Competence* goes further, in defining the appropriate delivery of those elements according to a prior definition. *Differentiation* is specifically relevant in both process and efficacy research, as it "refers to whether two or more treatments differ from each other along critical dimensions that are central to their execution" (Kazdin, 1994: 37). This has led to the idea

that certain behaviours are *prescribed* and others *proscribed*.

There are well-developed measures to assess the adherence of therapists across several types of therapy. This method, originally developed for the NIMH Treatment of Depression Collaborative Research Program (TDCRP; Elkin, 1994: 116), was called the Collaborative Study Psychotherapy Scale (CSPRS). Initially, it covered interpersonal therapy and cognitive therapy, as well as generic facilitating conditions and skills in clinical management. More recently the domains have been extended to include psychodynamic interpersonal therapy (Startup & Shapiro, 1993). Several other treatments have manuals to assess adherence and competence, and it has been recommended that all such scales be explicitly linked to a treatment manual (Waltz *et al*, 1993).

The examples from the TDCRP (Elkin, 1994) and Sheffield (UK) (Shapiro *et al*, 1994) studies demonstrate the conceptual areas covered (see Table 2). Each of the main therapeutic models considered covers a discrete domain, and some generic qualities are relevant across all types of therapy. More competent therapists (rated as such by their supervisors) were able to deviate appropriately from technical recommendations with more difficult patients (Rounsaville *et al*, 1988). There is also some evidence that patients of these therapists had better outcomes (O'Malley *et al*, 1988). Experienced clinicians tend to integrate elements from therapies which are differentiated in formal outcome studies. This calls into question one of the main paradigms of psychological treatments research: the so-called 'drug metaphor' implies that 'pure' or 'unadulterated' forms of treatment are likely to be most efficacious.

#### Clinical implications

- (i) The use of an explanatory rationale, structuring the session, and working towards focal agreed targets are hallmarks of modern structured brief approaches.
- (ii) The Vanderbilt II study suggested that the performance of therapists might actually be found to be worse if they adhere to a new treatment without detailed consideration of their psychological competence (Henry *et al*, 1994).

#### (d) Performance: synthesising adherence, competence and skilfulness

Competence scales have been rightly criticised in the past because they tended to reduce the complexity of therapy to a lowest common denominator. Recently, however, there has been increasing interest in higher levels of performance that could reasonably be called skilful. Milne *et al* (1999) have revised the competence scale for cognitive therapy. They distinguish six levels of competence. Interestingly, they take account of the literature on negative therapeutic effects and incorporate a 'Level 0' which represents a harmful state rather than a simple lack of competence, continuing through Novice, Advanced beginner, Competent, Proficient to Expert. It is interesting to note that their highest level of competence, Expert, is characterised as follows:

"The therapist no longer uses rules, guidelines or maxims. He/she has deep tacit understanding of the issues and is able to use novel problem-solving techniques. The skills are demonstrated even in the face of difficulties (e.g. excessive avoidance)."

Their suggestion for cognitive therapy suggests that at the highest level of competence there is an ability to transcend rules, guidelines and maxims, throwing into confusion any attempt to equate expert performance and adherence. Examples of the range of difficult situations across which competence needs to be demonstrated in *expert* practice might be an intensely suicidal patient, or a marked display of hostility.

In contrast to the research on treatment integrity (which draws heavily on the 'pure drug' metaphor), clinicians have turned increasingly to a hybrid approach where clinical skilfulness is measured in actual performance of the task. Hence, the paradigm suggested by Milne *et al* (1999) for cognitive therapy is likely to be of wider applicability.

#### Clinical implications

- (i) Skilfulness as a therapist can be conceptualised as the ability to adapt so as to stay broadly within a treatment paradigm even under unfavourable conditions.
- (ii) Skilled performance is more difficult to define and measure than adherence or competence.

**Table 2** Adherence and competence elements for interpersonal, cognitive, psychodynamic interpersonal, clinical management, facilitating and directive conditions

Interpersonal therapy (IPT)	Cognitive therapy (CT)	Psychodynamic interpersonal (PI) therapy <sup>1</sup>	Facilitating conditions and explicit directiveness conditions
<b>Interpersonal rationale</b>	<b>Cognitive rationale</b>	<b>Exploratory rationale</b>	<b>Facilitating conditions</b>
Interpersonal rationale	Relationship of thoughts and feelings	Exploratory therapy rationale	Supportiveness
Relating interpersonal change to therapy	Cognitive therapy rationale	Relating interpersonal change to therapy	Convey competence/expertise
	Relating improvement to cognitive change		Therapist's communication style
			Involvement
<b>Focus on feelings</b>	<b>Assessing cognitive processes</b>	<b>Shared understanding</b>	Warmth
Feelings about loss	Specific examples	Negotiating style	Rapport
Explore feelings	Reporting cognitions	Language of mutuality	Empathy
Acknowledge affect	Exploring personal meaning	Metaphor	Formality (–ve keyed)
Accept affect	Recognising cognitive errors	Disclosure	
	Identifying underlying assumptions	Understanding hypotheses	<b>Explicit directiveness conditions</b>
<b>Assessing interpersonal relationships</b>	<b>Evaluating and changing beliefs</b>	<b>Focus on 'here and now'</b>	Level of verbal activity
Interpersonal focus	Distancing of beliefs	Based on cues	Explicit guidance
Patterns in relationships	Examine available evidence-testing beliefs prospectively	Focusing	Subtle guidance (–ve keyed)
Communication patterns	Searching for alternative explanations	Confrontation	Receptive silence (–ve keyed)
Interpersonal expectations	Realistic consequences		
Relationship satisfaction	Functional value of beliefs	<b>Focus on feelings<sup>2</sup></b>	
	Practising 'rational responses'	Explore feelings*	
<b>Assisting change in interpersonal function</b>	<b>Behavioural focus</b>	Acknowledge affect*	
Changes desired in relationships	Planning/practising alternative behaviours	Acceptance of affect*	
Consider options for interpersonal change	Increasing pleasure and mastery	Limitations	
Ways of changing relationships	Scheduling/structuring activities	<b>Gaining insight</b>	
Developing new relationships	Self-monitoring	Patterns in relationships	
		Linking hypotheses	
		Explanatory hypotheses	
<b>Role transition</b>	<b>Homework</b>	<b>Structure</b>	
Focus on role transition	Homework reviewed	Sequencing interventions	
Exploring new role	Homework assigned	Structuring the session	
Evaluate loss in role change	Recording thoughts		
Preparation for new role			
<b>Interpersonal disputes</b>	<b>Collaborative structure</b>		
Focus on interpersonal disputes	Set and follow agenda		
Consequence of disparate expectations	Collaboration		
Relating disputed relationship to other relationships	Encourage independence		
Maintaining the dispute	Negotiate therapy contract		
	Explain therapist's direction		
	Summarising		
<b>Interpersonal deficits</b>			
Focus on interpersonal deficits			
Understanding past difficulties			
Using therapeutic relationship as a model			
Best aspects of previous relationships			

1. The PI scale is derived from Startup & Shapiro (1993), and all other scales from the rational sub-scales derived from the Collaborative Study Psychotherapy Scale (CSPRS) (Elkin et al, 1994).

2. The items marked \* are variants of the IPT scale items used in PI therapy.

### (e) Treatment definitions

The number of psychological treatments described far exceeds our capacity to differentiate them at all the levels defined above. All therapies could be reduced to a generic level: for example, distinguishing cognitive-behavioural, psychodynamic interpersonal, systemic and humanistic. However, there is still a substantial problem in allocating explicitly integrative models within the classification. A theoretical resolution of this would be the therapeutic equivalent of a colour chart. The proportion of each 'primary' therapeutic mode present could define a therapy (whether at the global level or at a session level).

An alternative approach has been to reduce the therapies under serious consideration to those which have well-defined methods and treatment manuals. 'Manualisation' has been another attempt to 'purify' the psychotherapy delivered, but

"it is foolish to believe that the use of manuals alone will 'standardise' a therapy. The actual delivery of therapy is dependent on the contributions and interactions that take place between . . . people" (Koss & Shiang, 1994: 675).

However, reviews of brief therapy suggest that manualised therapies have better outcomes than less formalised methods (Crits-Cristoph, 1992). Future directions in the development of manuals are likely to focus on broad definitions, with guidance about the underlying conceptualisation from which the treatment techniques would follow. There is likely to be a considerable delay before such integrated treatment manuals become common in both research and routine practice.

#### Clinical implications

- (i) Treatment manuals are often designed from the perspective of a research programme concerned to differentiate treatments, rather than of how best to give existing practitioners additional skills.
- (ii) There have been recent developments in the use of self-monitoring tools which focus attention on key therapeutic tasks (for example, in cognitive analytic therapy) (Bennett & Parry, 1998).

### (f) Therapeutic alliance

Most treatment methods and manuals now pay particular attention to one of the features of therapy that had previously been conceptualised as a 'common factor'. However, the therapeutic alliance has

emerged as the most consistent predictor of outcome across many studies in different models of psychotherapy (see Henry *et al*, 1994, for a review). Following Bordin (1979), the alliance has generally been construed as having three components: the therapeutic bond, agreement about the task and agreement about goals. The problem in terms of measurement has been that some aspects of 'alliance' could equally be seen as early or emerging outcomes. There are well-established and empirically validated scales to measure alliance, such as the California Psychotherapy Alliance Scales (CALPAS; Gaston, 1991), which minimise potential confounding with early outcome. However, there has been no research on whether the use of measures of alliance actually enhances routine clinical practice.

#### Clinical implications

- (i) Empirical measurement of the therapeutic alliance has not been used widely in clinical practice, other than in settings where it is combined with outcome measurement.
- (ii) Early warning of disruption to the therapeutic alliance can 'flag' the case for additional supervisory discussion.

### (g) Routine outcome measurement

For routine outcome assessment to become feasible, several conceptual and methodological issues need to be resolved. Surveys of outcome measures used in research and in routine practice show that very many measures are used (Mellor-Clark *et al*, 1997). Many are used in only one research project or service setting, which makes comparability extremely difficult. Many scales have poor referential data (particularly data applicable to service settings). Scales often have large numbers of items to increase reliability and to cover multiple domains. Scales have often not been checked for acceptability in different ethnic groups or validated across different languages. They often cover only a single perspective (for example, patient or informant).

Outcomes may be specific to a particular type of problem (for example, an eating disorder) or may refer to more general domains such as well-being, health, symptoms and function. The Compass System (Sperry *et al*, 1996) was developed as a systematic approach to assessing change in therapy, particularly in the context of health maintenance organisations

(HMOs). This approach suggested that outcome could be conceptualised in three phases: remoralisation, remediation and rehabilitation, covering improvement in well-being, symptoms and general life function respectively. The system also measured session-by-session alliance. It was possible to track patients' progress and flag cases needing closer attention by comparing session-by-session progress and variability against the main database.

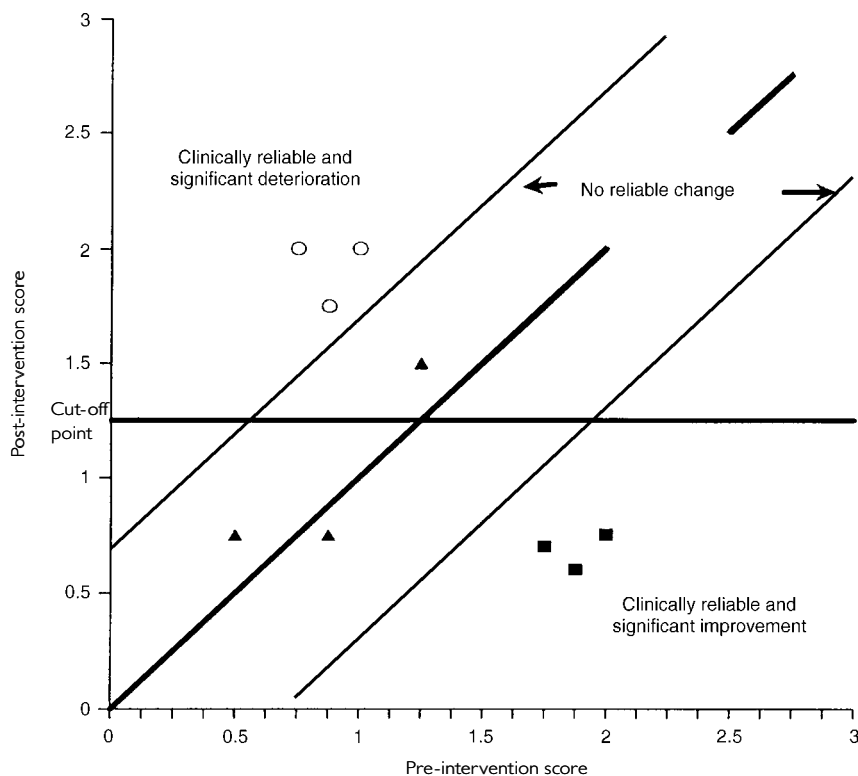
In the United Kingdom the CORE system (Core System Group, 1998) was developed from initial research about acceptability and from stakeholder views, in conjunction with several practice research networks (Barkham *et al*, 1998). The measures consist of a 34-item outcome scale (and two parallel short forms, for repeated use), an assessment measure and an end-of-phase summary. The measures can be scanned by computer or scored by hand for immediate use. The outcome measure is designed to assess well-being, problems or symptoms, and functioning, and has additional 'risk-flagging' items. The CORE system uses the measures to provide feedback to individual practitioners, teams and organisations using different modes of therapy and in a wide variety of settings (CORE System Group, 1998).

#### Reliable and clinically significant change

Measurement systems such as Compass and CORE, along with more specialised measures for particular problems, can then be utilised in an effectiveness framework using the concepts and methods of reliable and clinically significant change (RCSC). One of the main criticisms of efficacy research is that the results need to be interpreted at the group (or aggregated) level. Efforts to understand the outcome of an individual patient in the context of a whole service or outcome study have been limited by methodological and statistical problems (Evans *et al*, 1998).

Jacobson & Truax (1991) summarised a model of the measurement of change which reflected the complementary concepts of the statistical reliability of a change and its clinical significance.

Reliability reflects the extent to which an observed difference between scores (for instance, before and after treatment) is evidence of a true underlying difference in the patient. The statistic for estimating this reliability is the standard error of the difference score ( $s.e._{diff}$ ), which relates to the



**Fig. 1** Graphical plot of reliable and clinically significant change parameters (Evans *et al*, 1998) (reproduced by kind permission of BMJ Publishing Group).

standard deviation of the population and the reliability of the measure (Evans *et al*, 1998). If the change measured for an individual is more than 1.96 times the  $s.e._{diff}$  then such a change is unlikely to occur on more than 5% of occasions by chance. In practice this leads to a very simple way of representing change for a group of individuals on a two-dimensional graph, where the  $x$ -axis represents the pre-treatment score and the  $y$ -axis the post-treatment score on the same instrument. Every point on the graph can then represent an individual who has the corresponding pre-treatment and post-treatment scores. Figure 1 shows such a graph. The centre diagonal line represents all the points where there has been no change between before and after treatment ( $x=y$ ). The 'tramlines' on either side of the diagonal represent the limits of  $1.96 \times s.e._{diff}$  and so for anyone falling within the tramlines, a change could be attributed to chance. Those falling above the upper diagonal have reliably shown deterioration, whereas those below the lower diagonal line have reliably shown improvement.

#### Graphical representation of clinical change

This method of graphically representing change scores as single points in two-dimensional space (with inbuilt confidence intervals showing where change can be considered reliable) is an extremely helpful audit tool either for an individual clinician or to display the results for a whole department. Jacobson & Truax (1991) also drew up criteria to determine whether the change was clinically significant. There are many technical issues to be resolved about skewed distributions and the choice of cut-off points (see Evans *et al*, 1998), but essentially the argument indicates whether the individual has moved from a dysfunctional to a more functional sample of scores. This is a norm-referenced method and is, therefore, vulnerable to changes in the choice of reference parameters. Even so, it is arguably better than the alternative approach, which is simply to state an arbitrary drop in the measure used as the criterion for clinically significant improvement. There are now reasonably good cut-off

points for some measures (for example, the Hamilton Rating Scale: Grundy *et al*, 1996; Symptom Checklist-90-R: Tingey *et al*, 1996), but the methodology is still limited by the small number of normative samples for many instruments. A large normative dataset is needed to allow clinicians to focus on the expected outcomes for their particular individual patients.

#### Clinical implications

- (i) RCSC methods are useful to report progress for a whole department (or an individual's case-load) in a summary visual form.
- (ii) Normative data drawn from several referential samples are now available for commonly used instruments.

## DISCUSSION

### Practice research networks

We have given some examples of advances in measurement methods in psychotherapy. All are derived originally from research, and they also demonstrate possible applications of research in improving routine clinical practice. However, this would need an appropriate infrastructure. Practice research networks (PRNs; Zarin *et al*, 1996) provide such an infrastructure. These involve "... a large number of clinicians who agree to collaborate to collect and report data ... ." (Zarin *et al*, 1996: 147). The structure of a PRN helps to meet the dual agenda of evidence-based practice (EBP) and practice-based evidence (PBE). The PRN is typically linked with one or more academic centres which help to keep the group appraised of recent developments in the literature and disseminate recent systematic reviews.

### Use of large datasets

One of the advantages of a PRN in developing practice-based evidence is in generating very large datasets. This allows much better prediction at the level of the individual case.

A PRN is defined, somewhat tautologically, as a "network of clinicians that collaborate to conduct research to inform their day-to-day practice" (Audin *et al*, 2000). In contrast to most 'formal' research, PRNs utilise data gathered in 'real-world' practice settings rather than specifically orchestrated clinical trials, and large,

clinically representative, datasets can be developed.

Sperry *et al* (1996: 70–71) give a clinical example. They cite a 28-year-old female patient with a 2–3 month history of anxiety and depression. She complained of diffuse difficulties which were worse on workday mornings. The therapist provided five sessions of initial therapy, looking for underlying causes. The case manager expressed alarm at the review point because of evident deterioration in her clinical scores. After a second opinion, and a revised focus on the underlying work issues, the patient's clinical state improved.

### Case flagging

At the review point, when the case is 'flagged', it is possible to use more specific quality measures: Is the therapist following the procedures for this type of therapy? Is the formulation complete and of reasonable quality? Are there complicating factors (such as intercurrent substance misuse) which have been overlooked? The practitioner can then discuss the case formulation in a peer group to identify any factors which can be remedied.

In practice, the situation is not always as clear as this account suggests. There are two types of problem. Reducing the dataset to young, female, single patients with anxiety, depression and possible work-related problems will reduce the sample size dramatically, so that the confidence intervals increase (although Sperry *et al* (1996) demonstrate that reasonable confidence limits can be obtained). The second, and more profound, difficulty is that purely numerical case monitoring is likely to be impossible. Even the strongest advocates of case monitoring would only claim that these methods are adjuncts to clinical methods of supervision and case reviews.

### Good clinical practice: bridging the efficacy–effectiveness gap

Good clinical practice can be supported by drawing on various measurement perspectives. These will include

- recognition of the interventions being used and why they have been selected;
- ability to draw on different formulation methods, which are consistent with the chosen treatment method;
- recognition of threats to the therapeutic alliance and methods of repairing it

### CLINICAL IMPLICATIONS

- Use of reliable and clinically significant change methods can enhance clinical audit.
- Practice research networks help to evaluate effectiveness in clinical settings.
- Case formulation methods improve the accuracy of clinical interventions.

### LIMITATIONS

- The review is descriptive and not based on methods of systematic review.
- The use of very large datasets can produce statistical significance which is of little clinical relevance.
- The precision of cut-off scores in reliable and clinically significant change (RCSC) depends on the reliability of the measure used and the skewedness of the distribution.

FRANK R. MARGISON, FRCPSych, GRAEME McGRATH, FRCPSych, Department of Psychotherapy, Manchester Royal Infirmary; MICHAEL BARKHAM, PhD, JOHN MELLOR CLARK, MA, KERRY AUDIN, BSc, JANICE CONNELL, BSc, Psychological Therapies Research Centre, University of Leeds; CHRIS EVANS, MRCPsych, Rampton Hospital, Nottinghamshire and Tavistock Centre, London

Correspondence: Dr Frank R. Margison, Gaskell Psychotherapy Centre, Manchester Royal Infirmary, Sinton Grove, Manchester M13 0EU, UK. E-mail: frmargison@aol.com

(First received 5 March 1999, final revision 25 October 1999, accepted 10 November 1999)

across various challenging clinical situations;

- self-assessment and peer-review methods to evaluate the level of performance of the method in terms of adherence, competence and skilfulness;
- use of appropriate outcome measurement in routine practice, judged against relevant normative data, using reliable and clinically significant change methods;
- use of PRNs as an effective way of introducing practice-based evidence to complement the evidence-based practice paradigm.

### REFERENCES

- Audin, K., Mellor-Clark, J., Barkham, M., *et al* (2000) Practice research networks for effective psychological therapies. *Journal of Mental Health*, in press.
- Barkham, M. & Mellor-Clark, J. (2000) Rigour and relevance: practice-based evidence in the psychological therapies. In *Evidence-Based Health Care in Psychological Therapies*. London: Routledge, in press.
- , Evans, C., Margison, F., *et al* (1998) The rationale for developing and implementing core outcome batteries for routine use in service settings and psychotherapy outcome research. *Journal of Mental Health*, **7**, 35–47.
- Bennett, D. & Parry, G. (1998) The accuracy of reformulation in cognitive analytic therapy: a validation study. *Psychotherapy Research*, **8**, 84–103.
- Bordin, E. S. (1979) The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy Theory, Research and Practice*, **16**, 252–260.
- Cochrane, A. L. (1972) *Effectiveness and Efficiency. Random Reflections on Health Services*. London: Nuffield Provincial Hospitals Trust.
- Core System Group (1998) *CORE System (Information Management) Handbook*. Leeds: Core System Group.
- Crits-Cristoph, P. (1992) The efficacy of brief dynamic psychotherapy: A meta-analysis. *American Journal of Psychiatry*, **149**, 151–158.
- Department of Health (1996) *Psychotherapy Services in the NHS in England*. London: Her Majesty's Stationery Office.
- Elkin, I. (1994) The NIMH Treatment of Depression Collaborative Research Program. In *Handbook of Psychotherapy and Behaviour Change* (4th edn) (eds A. E. Bergin & S. L. Garfield), pp. 114–142. New York: Wiley.
- Elliott, R., Stiles, W. B., Mahrer, A. R., *et al* (1987) Primary therapist response modes: a comparison of six rating systems. *Journal of Consulting & Clinical Psychology*, **55**, 218–223.
- Evans, C., Margison, F. & Barkham, M. (1998) The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-based Mental Health*, **1**, 70–72.
- Gaston, L. (1991) Reliability and criterion-related validity of the California Psychotherapy Alliance Scales. *Psychological Assessment*, **3**, 68–74.

- Grundy, C. T., Lambert, M. J. & Grundy, E. M. (1996)** Assessing clinical significance: Application to the Hamilton Rating Scale for Depression. *Journal of Mental Health, 5*, 25–33.
- Henry, W. P., Strupp, H. H., Schacht, T. E., et al (1994)** Psychodynamic approaches. In *Handbook of Psychotherapy and Behaviour Change* (4th edn) (eds A. E. Bergin & S. L. Garfield), pp. 467–508. New York: Wiley.
- Hobson, R. P. (1993)** *Autism and the Development of Mind*. Hove: Erlbaum.
- Jacobson, N. S. & Truax, P. (1991)** Clinical significance: A statistical approach to defining meaningful change in psychotherapy. *Journal of Consulting & Clinical Psychology, 59*, 12–19.
- Kazdin, A. E. (1994)** Methodology, design and evaluation in psychotherapy research. In *Handbook of Psychotherapy and Behaviour Change* (4th edn) (eds A. E. Bergin & S. L. Garfield), pp. 19–71. New York: Wiley.
- Koss, M. P. & Shiang, J. (1994)** Research on brief psychotherapy. In *Handbook of Psychotherapy and Behaviour Change* (4th edn) (eds A. E. Bergin & S. L. Garfield), pp. 664–700. New York: Wiley.
- Luborsky, L. (1984)** *Principles of Psychoanalytic Psychotherapy: A Manual for Supportive-Expressive (SE) Treatment*. New York: Basic Books.
- Mace, C. & Margison, F. (1997)** Attachment and psychotherapy: an overview. *British Journal of Medical Psychology, 70*, 209–215.
- Margison, F. R., Loebl, R. & McGrath, G. (1998)** The Manchester experience: audit and psychotherapy services in northwest England. In *Rethinking Clinical Audit: The Case of Psychotherapy Services in the NHS* (eds R. Davenhill & M. Patrick), pp. 76–110. London: Routledge.
- Mellor-Clark, J., Barkham, M. & Evans, C. (1997)** *Talking Effectiveness: Providers' and Purchasers' Perspectives on Outcomes*. PTRCMemo 335. Leeds: Psychological Therapies Research Centre, University of Leeds.
- Milne, D. L., Baker, C., Blackburn, I. M., et al (1999)** Effectiveness of cognitive therapy training. *Journal of Behavior Therapy & Experimental Psychiatry, 30*, 81–92.
- O'Brien, T. B. & DeLongis, A. (1996)** The interactional context of problem-, emotion-, and relationship-focused coping: the role of the big five personality factors. *Journal of Personality, 64*, 775–813.
- O'Malley, S. S., Foley, S. H., Rounsaville, B. J., et al (1988)** Therapist competence and patient outcome in interpersonal therapy of depression. *Journal of Consulting & Clinical Psychology, 56*, 496–501.
- Omer, H. & Dar, R. (1992)** Changing trends in three decades of psychotherapy research: the flight from theory into pragmatics. *Journal of Consulting & Clinical Psychology, 60*, 88–93.
- Persons, J. (1989)** *Cognitive Therapy in Practice: the Case Formulation Approach*. New York: Norton.
- Roth, A. & Fonagy, P. (eds) (1996)** Translating research into practice. In *What works for whom? A Critical Review of Psychotherapy Research*, pp. 13–36. New York: Guilford.
- Rounsaville, B. J., O'Malley, S., Foley, S., et al (1988)** Role of manual-guided training in the conduct and efficacy of interpersonal psychotherapy for depression. *Journal of Consulting & Clinical Psychology, 56*, 681–688.
- Sackett, D. L., Richardson, W. S., Rosenberg, W., et al (1997)** *Evidence-based Medicine: How to Practise and Teach EBM*. London: Churchill Livingstone.
- Shapiro, D. A., Barkham, M., Rees, A., et al (1994)** Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic-interpersonal psychotherapy. *Journal of Consulting & Clinical Psychology, 62*, 522–534.
- , —, —, et al (1995) Decisions, decisions: Determining the effect of treatment method and duration on the outcome of psychotherapy for depression. In *Research Foundations for Psychotherapy Practice* (eds M. Aveline & D. A. Shapiro), pp. 151–174. Chichester: Wiley.
- Sperry, L., Brill, P. L., Howard, K. I., et al (1996)** Outpatient treatment outcomes. In *Treatment Outcomes in Psychotherapeutic and Psychiatric Interventions*, pp. 75–109. New York: Brunner/Mazel.
- Startup, M. & Shapiro, D. A. (1993)** Therapist treatment fidelity in prescriptive vs. exploratory psychotherapy. *British Journal of Clinical Psychology, 32*, 443–456.
- Stiles, W. B., Elliott, R., Llewelyn, S., et al (1988)** Assimilation of problematic experiences by clients in psychotherapy. *Psychotherapy, 27*, 411–420.
- Tingey, R. C., Lambert, M. J., Burlingame, G. M., et al (1996)** Assessing clinical significance: Proposed extensions to method. *Psychotherapy Research, 6*, 109–123.
- Waltz, J., Addis, M. E., Koerner, K., et al (1993)** Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting & Clinical Psychology, 61*, 620–630.
- Zarin, D. A., West, J. C., Pincus, H. A., et al (1996)** The American Psychiatric Association Practice Research Network. In *Outcomes in Assessment in Clinical Practice* (eds L. I. Sedderer & B. Dickey). Baltimore, MD: Williams & Wilkins.