

METHOD PAPER

## Developing practice-based evidence: Benefits, challenges, and tensions

ROLF HOLMQVIST<sup>1</sup>, BJÖRN PHILIPS<sup>1</sup>, & MICHAEL BARKHAM<sup>2</sup>

<sup>1</sup>Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden & <sup>2</sup>Centre for Psychological Services Research, University of Sheffield, Sheffield, UK

(Received 8 April 2013; revised 10 October 2013; accepted 26 October 2013)

### Abstract

Attempts to regulate service delivery in line with results from randomized trials have been vigorously debated. In this paper, results from practice-based studies using the CORE System illustrate the potential to enrich knowledge about the actual outcome of psychological therapy in routine care. These studies also provide data for important questions in psychotherapy research, like orientation differences, the importance of the therapist factor, number of sessions needed for clinical effect, and the alliance–outcome question. Obstacles and challenges in making such studies are illustrated. In conclusion, arguments are put forward for introducing a common measurement system that strikes a balance between clinicians' questions and the need for comparable data, and that encompasses the complexities of patients' reasons for seeking psychological help.

**Keywords:** outcome research; mental health services research

### Introduction

Recent decades have seen an immense increase in developing and delivering improvements in accessing quality psychological healthcare for people in need. Efforts to enhance the effectiveness of psychological treatments have tended to follow one of two possible tracks. One track has involved developing psychological treatment methods “in the lab,” testing their efficacy in well-controlled trials, and then disseminating them to routine care (Hunsley & Lee, 2007). This paradigm is often termed *evidence-based practice* (McHugh & Barlow, 2012). Considerable effort and money in the UK and in Sweden have been invested in disseminating such treatments to service providers and education programs. By contrast, the other track has been to study the effects of psychological treatments as they are delivered in routine clinical practice, a paradigm termed *practice-based evidence*<sup>1</sup> (Barkham, Hardy, & Mellor-Clark, 2010; Barkham & Margison, 2007). In such studies, allowance is made for variation in patient symptoms and comorbidity, therapist competence, and treatment

contexts. These components can be investigated within the naturally occurring hierarchical nature of large-scale practice-based datasets, which allows for the use of sophisticated multilevel modeling techniques. And then findings from practice-based studies can later be tested in systematic treatment trials that, in turn, lead to further practice-based studies, thereby resulting in a naturally occurring cycle of research activity (Barkham & Mellor-Clark, 2003).

These two tracks are sometimes—mistakenly—viewed as competing with each other. However, they are in fact complementary paradigms, as both are needed in order to build a robust and rigorous science of the psychological therapies (Barkham, Stiles, Lambert, & Mellor-Clark, 2010). The purpose of the present article is to set out both the common and contrasting obstacles and tensions experienced in two countries—Sweden and the UK—in pursuit of redressing the balance between evidence-based practice and practice-based evidence.

At the health political level, there are clear similarities between the UK and Sweden. The government and state authorities have large influence on the

---

Correspondence concerning this article should be addressed to Rolf Holmqvist, Linköping University, Department of Behavioural Sciences and Learning, Linköping, 58183 Sweden. Email: [rolf.holmqvist@liu.se](mailto:rolf.holmqvist@liu.se)

health service in both countries. Treatment of psychiatric disorders and psychological problems is to a large extent delivered by publicly financed service units (e.g., primary care centers and psychiatric outpatient clinics). The governments in both countries have taken strong initiatives to promote evidence-based treatments in these treatment services, requiring that only therapists with knowledge of evidence-based methods should provide treatment.

There are at present also some unifying components between Sweden and UK with regard to psychotherapy research and practice. In ongoing programs of practice-based studies in both countries, predominantly, although not exclusively, the same measurement system is being used. In both countries, there is also a dominance of the paradigm of evidence-based practice, best exemplified by the randomized controlled trial.

Accordingly, in this article, first we set out the origins and philosophy of our common measurement system. Second, we focus on the role of such a measurement system in underpinning the paradigm of practice-based evidence and the need for studies in this paradigm. And third, we draw out benefits, challenges, and tensions relating to implementing practice-based studies in Sweden and the UK together with the generic issues associated with developing a common outcome measure for practitioners.

### The CORE Measures and System: Origins and Philosophy

The CORE program is a fluid and organic portfolio of work that began in the mid-1990s with the aim of developing a free-to-use, pantheoretical outcome measure which tapped the core components that practitioners considered as central when evaluating patient change: subjective wellbeing, problems, functioning, and risk. Accordingly, the measure—and much of the subsequent work—has been premised on meeting practitioners' needs and building an evidence base that speaks to local service requirements but that also can be the basis for building a cumulative evidence base to complement trials-based evidence.

The CORE initiative arose in response to the SPR-initiated symposium revisiting the concept of a core outcome battery (Strupp, Horowitz, & Lambert, 1997), which had its own origins in a seminal symposium held in 1970 focusing on the measurement of change (Waskow & Parloff, 1975). These two symposia provided the impetus for an initiative in the UK whereby a UK charity—the Mental Health Foundation—funded two successive pieces of research focusing first on the acceptability and feasibility of developing a core outcome battery and then on the development of a core outcome

measure. This work yielded the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM; Barkham et al., 2001; Evans et al., 2002). A further and final grant supported the implementation of the CORE-OM (i.e., a roll out to services), which also comprised a wider CORE system incorporating the collection of contextual information within which patient outcomes could be placed (for details, see Mellor-Clark & Barkham, 2012). These data—outcomes and contextual—were supported initially by a personal computer version of the CORE system (PC-CORE) and more recently by a web-based support—CORE Net. This information technology (IT) platform provides services with the facility for managing their data using multiple options and filters that determine the level of complexity in data reporting required by a service.

The original CORE-OM became the parent version for a family of CORE measures that derived from the original items (although with some adaptations), which yielded the following variants (see Barkham, Mellor-Clark, et al., 2010, for summary details): (i) two parallel shorter versions for intensive research (CORE-Short Form, A and B), (ii) a 10-item version (CORE-10), (iii) a five-item monitoring version (CORE-5), as well as versions for (iv) the general population (GP-CORE), (v) young people (YP-CORE), and (vi) people experiencing learning difficulties (CORE-LD). In addition, subsequent research has yielded the identification of six items from the CORE-OM to determine health utility status (CORE-6D; Mavranouzouli, Brazier, Rowan, & Barkham, 2013; Mavranouzouli, Brazier, Young, & Barkham, 2011). Most importantly, and crucial to the potential of building collaboration with practitioners, there has been an ongoing program of securing standardized translations of the CORE-OM and its variants into more than 25 languages, including a Swedish version (Elfström et al., 2013).<sup>2</sup>

The development of a comprehensive option of measures based on the CORE-OM together with the contextual information relating to patients' presenting problems and service provision provides a *measurement system*—as opposed to a simple outcome measure—that provides the core component for delivering on the agenda of practice-based evidence. In turn, practice-based evidence provides the overarching framework for collaborating with practitioners with the aim of both improving practice and building an evidence base to complement trials methodology.

### Practice-Based Evidence in the UK and Sweden

Although a bona fide measurement system is the key component for practice-based evidence, other

components are also required. These include a commitment to the philosophy of practice-based evidence, an infrastructure to support the research activities, and a willingness (and technical ability) to share data. Perhaps a key axiom to the philosophy underpinning practice-based evidence is that practitioners have a genuine sense of local ownership regarding the measurement system and the data it yields, be it in relation to an individual practitioner or at the level of psychological services. Accordingly, in this paradigm, the accumulation of evidence assumes a *bottom-up model* whereby routine data are used at an individual level and locally within the service but then also accumulated across services and used to generate a higher-order evidence base.

However, such a paradigm runs counter to the dominant model of evidence espoused in many countries. For example, practice-based studies carried out in Sweden and the UK have been implemented in national contexts that place the primary value on research methods (and interventions arising from such studies) endorsed by national bodies such as the National Institute for Health and Clinical Excellence (NICE) in the UK and the Socialstyrelsen (the National Board for Health and Welfare) and the Swedish Council on Technology Assessment (SBU) in Sweden. These national bodies place a central value on a *top-down model* in which evidence from randomized trials is then espoused in routine practice via clinical guidelines. Adherence to these guidelines at the level of individual services can then be evaluated by national audit—in the UK this has been achieved through the National Audit of Psychological Therapies involving feedback to services in the form of benchmarking their compliance against overall standards (Royal College of Psychiatrists, 2011). Similarly in Sweden, governmental authorities (e.g., the Ministry of Health and Social Affairs) base resource allocation to health services that provide specified evidence-based treatments.

Set against the dominant model of evidence-based practice, research programs utilizing routine outcome data have been carried forward. In the UK, this program of work has considered a range of issues including the benchmarking of services (Barkham et al., 2001; Evans, Connell, Barkham, Marshall, & Mellor-Clark, 2003), the comparative effectiveness of different models of therapies (Stiles, Barkham, Mellor-Clark, & Connell, 2008; Stiles, Barkham, Twigg, Mellor-Clark, & Cooper, 2006), and enhancing our understanding of dose-effect relations and the role of the good-enough model in providing an account of the dose-effect curve (Barkham et al., 2006). More recently attention has focused on the area of therapist variability (Saxon & Barkham, 2012). Research has also focused on

patient change via work addressing the phenomenon of sudden gains (Stiles et al., 2003), responsiveness (Stiles, Barkham, Connell, & Mellor-Clark, 2008), as well as the prediction of patient change (Lutz et al., 2005; Stulz, Lutz, Leach, Lucock, & Barkham, 2007).

In Sweden, a research program comprising data from psychological treatment in primary care and psychiatric clinics has recently started. To date, analyses have been carried out on treatment outcome with different treatment methods (Holmqvist, Ström, & Foldemo, 2013), on different aspects of the relationship between treatment alliance and outcome (Falkenström, Granström & Holmqvist, 2013a, 2013b; Larsson, Andersson & Holmqvist, 2013; Sinclair & Holmqvist, 2013), and on the significance of congruence between patient and therapist in the perception of symptoms (Holmqvist, Philips, & Mellor-Clark, 2013). Collectively, these research activities illustrate how the practice-based paradigm focuses on factors such as patient resources, therapist and relationship qualities, and setting variables.

A key advantage of the practice-based paradigm is that it provides possibilities to test treatment methods that clinical experience has supported and to understand patient and treatment moderating factors that influence results in clinical practice (Cahill, Barkham, & Stiles, 2010). An apparent advantage with practice-based studies is that they may elucidate the significance of contextual factors for the results. It is important to analyze what conditions may determine or influence outcome results in routine practice. Recent examinations have found large variations in outcome scores depending on which subsample of patients is studied and what outcome measures are used, with effect sizes ranging from .60 to 1.95 (Barkham, Stiles, Connell, & Mellor-Clark, 2012).

### Benefits of Practice-Based Studies

Although the twin framing of evidence-based practice and practice-based evidence is appealing as an overarching model for broadening our knowledge base about the psychological therapies, it is important to have evidence of the added value—the additional benefits—of the latter approach. Two benefits of practice-based evidence are (i) local ownership by practitioners and services, and (ii) cumulatively enhancing evidence in areas in which trial designs are not appropriate. Here we provide examples of these two yields.

#### Local Ownership

In contrast to the adoption of outcome measures via government mandate, a practice-based approach

would aim for a service to select a measure, from a battery of bona fide measures, that best suited their purpose. Exercising choice in selecting a validated measure would greatly facilitate a sense of ownership of the resulting data. Two UK examples where services have adopted CORE and provided accounts of its implementation are the Women's Counselling and Therapy Service (Leeds; [womenstherapyleeds.org.uk/wp-content/.../What-Happened-Next.pdf](http://womenstherapyleeds.org.uk/wp-content/.../What-Happened-Next.pdf)) and My Sister's Place (Middlesbrough; [www.coreims.co.uk/site\\_downloads/MSP\\_study\\_of\\_excellence.pdf](http://www.coreims.co.uk/site_downloads/MSP_study_of_excellence.pdf)). Both services have been set up for women, the former to provide long-term therapy and the latter as a response to domestic violence. Both adopted CORE around 2004–2005 and have integrated it into their service, and both have been recipients of awards from the British Association for Counselling and Psychotherapy in recognition of their work: the Women's Counselling and Therapy Service received the award for Outstanding Research Project (2011) while My Sister's Place received the award for Excellence in Counselling and Psychotherapy (2011).

In recounting the adoption of CORE, My Sister's Place have reported that, given the main goal was to improve service quality for clients, CORE “was enthusiastically embraced by all members of the counselling team.” Adopting CORE was facilitated by therapists being given as much support as necessary to ensure they were using the system effectively, including one-to-one training sessions. And for practitioners, CORE is presented as a means for them to reflect on their practice and develop as skilled practitioners rather than being used as a means of assessing their performance. The impact of adopting CORE included the following: Improved assessment procedures; development of a domestic violence-specific therapeutic model; support of additional therapeutic provision: alerting to an increase in the number of clients presenting with more complex trauma; building working partnerships with external referral sources and promoting the service to other referrers; and demographic data have been used to improve therapeutic provision to more marginal groups (e.g., a Disabled Outreach Therapy Service has been established, which offers therapy for disabled clients in their homes). Overall, it enabled the service to adapt to the changing needs of the clients, to expand the service, and to increase service provision and provide an evidence base to support funding applications for the service.

### Enhancing Evidence

One of the central issues within the field of psychotherapy research that can best be addressed with the broad collection of data that practice-based studies

make possible is the phenomenon of natural variability in aspects of human performance and service delivery. By definition, trials place constraints on a range of features in order to test a candidate intervention. By contrast, practice-based studies have the potential for understanding natural variability when not constrained by strict treatment protocol. Two examples of this would be therapist *drift* (e.g., Waller, 2009) and therapist *effects* (e.g., Baldwin & Imel, 2013). The former arises from therapists in routine practice not implementing proven techniques in keeping with the theoretical model of therapy delivered, and the latter captures the phenomenon that not all therapists are equally effective—that is, some yield consistently better outcomes. Both these examples place a focus on the practitioner rather than the intervention per se and both are features of routine practice.

In relation to therapist effects, a recent review of this field analyzed 46 studies that used random effects (Baldwin & Imel, 2013). While the overall effect was 5%, there was a significant difference in the extent of therapist effects between efficacy (trials) and effectiveness (naturalistic) studies, with a mean therapist effect for trials of 3% while for naturalistic studies it was 7%. Hence, while the overall (average) effect is in the region of 5%—similar to that of the therapeutic alliance—practice-based studies reveal the effect to be larger. One conclusion that the authors arrive at is “Advancement in our understanding of therapist effects will likely only come from larger studies of therapist effects .... Researchers will likely need to collaborate and pool resources to obtain sufficient amounts of data. It is a challenge but a challenge we must face if we are to move this literature forward.” (p. 280).

One example of this collaboration in aid of yielding larger studies is a report published subsequent to the above review and using CORE-OM data derived from 22 services and comprising one of the largest studies of therapist effects, in which 119 therapists each saw a minimum of 30 patients (Saxon & Barkham, 2012). After adjusting for case-mix, the therapist effect was 6.6%. However, of particular note was the finding that when considering the CORE scores (but excluding the Risk component), a low score yielded a therapist effect in the region of 1–2%. However, as patient intake scores increased, the therapist effect rose upwards to 10%. That is, the higher the patient initial severity score, the more it matters whom the patient sees. This makes clinical sense but is a finding that is unlikely to have arisen from an RCT. Although practice-based studies provide a context within which to investigate these effects, we need to be able to combine process and

qualitative research components in order to be able to unpack therapist effects.

More generally, in comparison with the strict inclusion criteria in efficacy studies, the results from practice-based studies underline the satisfactory results of routine psychotherapy treatment. They emphasize the importance of testing routine treatments in order to understand their effectiveness. But practice-based studies not only contribute to a better understanding of treatment results in routine care, they also open up for analyses issues that often cannot be properly handled in randomized studies considering their restriction of range in key variables.

### Key Foci in Practice-Based Studies

In this section, we set out briefly some of the arguments for practice-based studies, focusing on the key areas relating to patients, therapists, and services providing psychological therapies.

**Patients.** Patients have complex combinations of psychiatric, relational, and social problems that the treatment must handle. Stirman, DeRubeis, Crits-Christoph, and Brody (2003) found that results from RCTs performed in research settings may be less useful in clinical practice as the patients have heterogeneous psychopathology. More than half of community outpatients had primary diagnoses that were not represented in the existing psychotherapy outcome literature.

**Therapists.** Therapists often have heterogeneous therapeutic competencies. Many therapists are trained in several treatment methods, and it is common for therapists to describe themselves as integrative (Larsson, Kaldo, & Broberg, 2009). It is probable that therapists often combine treatment approaches according to their clinical sensibility and experience (Street, Niederehe, & Lebowitz, 2000). It is also probable that services vary in the extent of existing competencies among their staff and in their encouragement of developing diverse competencies.

**Treatment length.** Treatment length varies depending on symptom reduction rate and extraneous factors and not according to a manual (Baldwin, Berkeljon, Atkins, Olsen, & Nielsen, 2009). Findings from randomized studies with manualized treatments do not usually give information about number of sessions needed for symptom improvement, as guidelines usually prescribe treatment length. However, in actual practice treatment duration may differ considerably from the manuals' recommendations. Previous studies have suggested a general asymptotic

dose-response rate for improvement (Howard, Kopta, Krause, & Orlinsky, 1986), but a linear "good-enough" model has recently been proposed (Baldwin et al., 2009; Barkham et al., 2006), implying that the individual symptom reduction trajectory determines when the therapy will terminate.

**Under-represented treatments.** Some treatments are rarely or never studied in systematic trials although they represent a large part of actual therapeutic work in routine psychiatry (Pingitore, Schefler, Schwalm, Zarin, & West, 2001; Tanielian, Marcus, Suarez, A. P., & Pincus, 2001). In contrast to manual-based methods, some of which have been repeatedly studied in efficacy trials, these treatments have hardly ever been studied except as comparison treatments, often under disadvantageous conditions. Several studies have indicated that the effects of treatment methods do not differ in practice studies (Stiles, Barkham, Connell et al., 2008; Holmqvist, Ström et al., 2013) but more studies on frequently used methods are needed. This question includes the issue of interaction between service unit and treatment method. Knowledge about this apart from clinical lore is virtually non-existent. An example is supportive treatment, which has hardly ever been tried in an RCT as a treatment purported to be effective (Budge, Baardseth, Wampold, & Fluckiger, 2010). Although supportive therapy has been described in several handbooks (Novalis, Rojcewicz, & Peele, 1993; Pinsker, 1997; Winston, Rosenthal, & Pinsker, 2012), it is often described in pejorative terms (Hellerstein & Markowitz, 2008).

**Therapist effects and training.** As noted above, particularly interesting is the possibility of detecting therapist effects. Although the size of therapist effects has been debated (e.g., Elkin, Falconnier, Martinovitch, & Mahoney, 2006; Kim, Wampold, & Bolt, 2006), which factors contribute to them is not well understood (Crits-Christoph et al., 1991; Luborsky, McLellan, Diguier, Woody, & Seligman, 1997; Wampold & Brown, 2005). It is also important to understand the potential interaction between therapist and treatment method. One question is: What is the level and extent of therapist training that contributes to outcome? This issue has been vigorously debated but has delivered meager results (O'Donovan, Bain, & Dyck, 2005). The few studies that have been done often suggest that training has limited importance. This issue is too often evaded in research. In contrast to RCTs in controlled settings, practice-based studies include staff with various training levels and make it possible to assess the importance of psychotherapy training and experience.

**Patient-treatment matching.** In a natural clinical setting, efforts are often made to match each patient to a suitable treatment, unlike the randomization procedure used in clinical trials. Both formal assessment procedures might be used or more informal clinical judgments. Hence, the ordinary routine in a clinic often includes assigning patients with different types of problems and different treatment goals to the therapy methods available at the clinic (Philips, 2009). A naturalistic study design appears to be the best way to measure therapy process and outcome under such clinical conditions.

**Dropouts.** Outcome in treatment studies is usually measured as symptom reduction. A large problem arises, however, in the guise of dropouts from treatment (Baekeland & Lundwall, 1975; Clarkin, Levy, Lenzenweger, & Kernberg, 2004; Klein, Stone, Hicks, & Pritchard, 2003; Reis & Brown, 2006; Wierzbicki & Pekarik, 1993; Young, Grusky, Jordan, & Belin, 2000). It is striking that figures on treatment dropout vary widely between studies (Barrett, Chua, Crits-Christoph, Gibbons, & Thompson, 2008; Chiesa, Wright, & Neeld, 2003; Hoffman, 1985; Kazdin & Mazurick, 1994; McMurran, Huband, & Overton, 2010). The overall mean for dropout from treatment in routine practice seems to be around 50%. This is, of course, a large problem both for the individual, for the service, and for society at large. It is therefore also important to measure outcome as the proportion of patients completing the treatment course.

**Service effects.** Differences in patient outcome between delivery services have been noted in some studies (Brown, Dreis & Nace, 1999; O'Brien, Fahmy, & Singh, 2009). In randomized studies, attempts are made to reduce systematic variation in all variables except between the treatments—the independent variable. Results are considered to be generalizable to all services, across cities or regions or even countries. In practice-based studies, allowance is made for and interest is focused on differences in the treatment deliverance context. In practice, it is hard to imagine a randomized study where patients are randomized to different treatment services. It is, however, quite conceivable that services differ in the quality and extent of delivery of treatment (Bickman, 2008; Bickman, Kelley, Breda, Vides de Andrade, & Riemer, 2011), and it is important to understand better the factors that contribute to such potential differences, which may be due to different treatment models but also to differences in organizational structures. The significance of organizational factors for patient outcome has only received limited research interest.

## Benefits for Clinicians

Practice-based studies require close cooperation with the participating clinicians. The question of “What’s in it for me” must be taken seriously. As an example, an ongoing study in both the UK and Sweden explores clinicians’ thoughts about what kind of feedback they want. Using a randomized design, the influence of different therapist attitudes to sessional feedback on the factual use of feedback and on outcome is explored.

## Challenges in Practice-Based Studies

As an example of challenges that have to be met in practice-based data collection, we offer here some experiences from an ongoing project in Sweden. This study was initiated at the request of therapists in primary care who feared losing their jobs as a consequence of organizational changes. They were strongly motivated to participate and to engage patients, and data from about 1200 patients were obtained. The design was essentially an outcome study with measurements at each session, using the CORE-OM and the Working Alliance Inventory (Hatcher & Gillaspay, 2006). Psychiatric clinics in the same region decided to use the same design in order to make comparisons with the primary care data possible. At this stage, more hurdles and problems occurred. Although service management was positive, the therapists were more ambivalent. Common reasons for hesitancy were that the administration and completion of questionnaires took too much time, that patients were referred to neuropsychiatric examinations rather than to psychotherapy and thus not possible to include, and that many patient contacts could not really be considered as therapies. Still another reason may have been that therapists were not certain about what kind of feedback they would obtain.

Patient participation does not, however, seem to be a problem. Although some patients decline to participate, most of them accept and a few express appreciation about the study. Data attrition from patients is low. At present, the percentage of delivered CORE-OM forms, in relation to the number of sessions that the therapists have reported that patients have attended, is about 97%.

At present, two strategies are used in order to engage therapists. The first is to accept that not all therapists at a specific service participate with their patients. At these services, information about service performance will not be available unless more therapists choose to participate. The other strategy, adopted by some clinic managements, is to mandate that participation by all therapists is compulsory.

Nine agencies have contributed data to the study so far and shortly another 15 agencies will join the study. Some preliminary analyses have been made. One study found that supportive treatment, the most frequently used treatment orientation, compared well with more systematic treatments when adequate control for relevant factors such as patient age and treatment length was made (Holmqvist, Ström et al., 2013). Other studies have found that treatment alliance predicts outcome from the first session onwards, both from session to session and for the whole treatment, even when outcome for each session has been controlled (Falkenström et al., 2013a, 2013b), and that sessions where the alliance has a rupture-resolution pattern have better outcomes than “stable” sessions (Larsson, Andersson, & Holmqvist, 2013). In still another analysis of the data, it was found that congruence between patient and therapist in the perception of the patient’s level of depression and anxiety did not influence either treatment alliance or treatment outcome (Holmqvist, Philips et al., 2013).

Feedback is a central part of practice-based studies. In this study, feedback has, up to now, been given to those therapists with a substantial number of patients in the form of mean CORE-OM scores for the whole scale and subscales at the start of therapy, and as mean change scores for patients who have completed the treatment. Therapists are also informed about mean scores on the WAI ratings, on several social and psychiatric characteristics of the patients, and about treatment process variables.

As a part of this project, a special study of the effects of session-wise feedback, using CORE Net, is being carried out at two outpatient psychiatric clinics. The aim of this study is primarily to capture the reactions of therapists to the possibility of getting immediate information about a patient’s symptoms and problems. It is not self-evident that immediate feedback, when offered, is always wanted or regularly used. Data suggest that it is primarily when the patient’s condition worsens without the therapist being aware of it that session-wise feedback substantially increases outcome (Lambert, 2010).

An interesting aspect of studies collecting large data sets at different services is that such designs may allow for meaningful comparisons between units. In Sweden, this is a very timely issue, as a radical change in the provision of medical and psychiatric care has occurred during the last 15 years. From a system that almost entirely relied on public provision of health care, Sweden now has opened for private, even profit-driven, providers. But still, the main part of the care is paid from taxes. Thus, the state, counties, or cities allow public and private agencies and companies to compete for the right to deliver

care. This competitive situation has come about rapidly, and authorities often have an underdeveloped ability to negotiate contracts. Input variables, like opening hours and number of posts, are often used in the contract negotiations instead of outcome figures. Recently, researchers and agency representatives have started discussing better ways of handling the competitive situation. The CORE project might give important knowledge for the further development of good care for less money. Several authors (e.g., Lambert, 2010) have noted the surprising fact that both researchers and care providers have been content “to use informal judgments instead of patient outcome data to evaluate therapists” (p. 175). The same verdict pertains to the service level.

### Other Challenges

Evaluations of treatment results at therapist or service level have a value for society at large. But there is a risk that both public agencies and private companies may want to present more positive figures of treatment outcome than the data really warrant. An important challenge in studies at the service level is to publish figures with caution, always analyzing therapist and service effects with a caseload perspective and putting results within the frame of the socio-economic context.

### Tensions

In this final section, we focus on methodological issues pertaining to measurement within practice-based evidence and the associated tensions. Our rationale is that a key axiom of practice-based evidence is the rigor, robustness, and relevance of the measurement system that lies at the heart of research efforts. This is a major foundation for enabling practitioners to adopt and receive appropriate acknowledgement as a contribution to the knowledge base for the psychological therapies.

### Outcome Measures: Being Fit for Purpose and Fitting the Political Space

A major tension at the level of measure development is to devise a measure that is informed and/or endorsed by practitioners and service users—that is, ensuring measures are fit for purpose and appropriate for specific populations—whilst at the same time delivering to a *national agenda* in an increasingly outcome-oriented and politicized health delivery system—that is, meeting the requirement of the current political space. The dominant focus on diagnosis has resulted in an over-valuing of

symptom-specific (i.e., unifocal) outcome measures. This situation is exemplified in the UK, where the PHQ-9 (Kroenke, Spitzer, & Williams, 2001) and GAD-7 (Spitzer, Kroenke, Williams, & Löwe, 2006) have been mandated for use in the Improving Access to Psychological Therapies (IAPT) initiative and, accordingly, have become dominant. These two measures are derived directly from the DSM-IV criteria for depression and anxiety respectively. In contrast to these symptom-specific measures, the dimensions comprising the CORE-OM were based on surveying practitioners' views as to what domains of change were important for their patients (Barkham et al., 2001).

An intrinsic strength of the CORE-OM is its coverage of a broad scope of symptoms as well as well-being and aspects of social and general functioning (including close relationships) together with a robust coverage of risk (to self and to others). Although the IAPT initiative also uses a short measure of functioning—the Work and Social Adjustment Scale (Mundt, Marks, Shear, & Greist, 2002)—definitions of *moving to recovery* are made using only the PHQ-9 and GAD-7. Hence, functioning does not contribute to the definition of recovery, which is operationalized as a patient's score on both PHQ-9 and GAD-7 measures falling below measure-specific cut-off scores.

While the language of symptoms and diagnoses is a good fit within trials methodology, the language of functioning is probably a fairly dominant index for everyday life. It may be that future measure development will focus more on the multiple aspects of functioning and such instruments may become better indicators of meaningful patient change. Until then, a conceptual problem remains in assuming that combining independent symptom-based measures represents an integrated clinical sense of recovery for a patient and that what matters most to a patient is achieving a specific score on wholly symptom-based measures. Indeed, a subsequent independent evaluation by service users of 24 commonly used outcome measures rated the CORE-OM more positively than the PHQ-9 (Crawford et al., 2011). Interestingly, the measure rated most highly by service users was the Warwick-Edinburgh Wellbeing Scale, perhaps because all the items were framed positively (Tennant et al., 2009). Certainly, any future measure development needs to ensure it is informed by the views and aspirations of service users.

A recent development within the CORE methodology is the ability to derive a health utility state from the CORE-OM via the CORE-6D (Mavranouzouli et al., 2011; Mavranouzouli et al., 2011, 2013), a facet that is not available with current symptom-

based measures and in which the EQ-5D (EuroQol Group, 1990) has been prominent. However, the ability of a generic measure that has shorter versions for repeated administration together with the ability to generate a health utility state is a potentially rich measurement source.

### Outcome Measures: Empirical and Conceptual Mapping

Notwithstanding the strengths of the CORE measure, future research effort needs to progress the *conceptual and empirical mapping* between outcome measures—sometimes referred to as the *walk across*—whereby practitioners have choices from a selection of bona fide measures but with each measure being able to be mapped onto a uniform or standard benchmark denoting valued change. For example, earlier work produced walk-across tables for use by researcher and practitioners between the CORE-OM and BDI-I (Leach et al., 2006) and subsequent work compared the PHQ-9 with the CORE-OM (Gilbody, Richards, & Barkham, 2007).

Thus, although we have described advantages with the CORE system, it is important to realize the tension arising from any specific measure having dominance, as it may freeze the scientific field and reduce practitioner and—by implication—patient/service user choice. Simple logic would suggest that no outcome measure is immune to future measure development. Hence, service delivery systems and national government policies need to be able to adapt and adopt a range of bona fide measures that meet a given set of required criteria in order that practitioners can select the measure that best suits their needs and that the empirical relationship between these measures is known. Moreover, collaboration with practitioners is best achieved if the measurement system operates at a holistic level akin to the practitioner working with the whole person of the patient rather than with fragmented parts determined by diagnosis and dominated by symptom-specific measures.

A key process that needs to be achieved in practice-based evidence is to ensure a sense of local ownership by practitioners in the data they collect. Part of this equation is that the measure itself is consistent with the theoretical models used by practitioners and acceptable to patients. Addressing this first step is one move towards responding to the *colander effect* in which data collected every day across clinical services are lost (i.e., slip through the holes of the colander) through lack of infrastructure and coordinated effort (Kazdin, 2008).

## Some Concluding Words

Practice-based studies of psychological treatment in routine care offer potentially rich sources of knowledge that complement the knowledge gained from randomized trials. The variability in patients, therapists, and settings mirrors the forms of treatment that patients in routine care are offered. At this stage in the development of practice-based evidence, the challenge is to consider and decide, on a common basis both within and between countries, as to the kind of information that should be collected that will enrich and stimulate treatment providers, therapists, patients, and therapy researchers. The day may have come for decisions on common measurement methods that were initiated several decades ago.

## Acknowledgments

The CORE initiative in the UK has been a collaborative venture and reflects the ideas and efforts of many practitioners and patients as well as the CORE System Trustees (MB, Chris Evans, Richard Evans, & Frank Margison) and CORE Information Management Systems (John Mellor-Clark & Alex Curtis Jenkins). The recently launched CORE-based practice network in Sweden is growing due to the collaborative effort of researchers, patients, therapists, and service managers. We want particularly to thank Tommy Skjulsvik and Ylva Gidhagen Söderberg.

## Notes

- <sup>1</sup> This paradigm has also been termed *practice-oriented research* (Castonguay, Barkham, Lutz, & McAleavy, 2013).
- <sup>2</sup> This research program has been led by Professor Chris Evans and enquiries should be directed to [chris@psyctc.org](mailto:chris@psyctc.org). Translations are available in the following languages: British Sign Language (BSL), Brazilian, Portuguese, Argentinian, Spanish, Romanian, Bulgarian, Arabic, French, Czech, Bangla/Bengali/Sylheti, isiXhosa, Catalan, Farsi, Kurdish, Kannada, Hindi, Urdu, Tamil, Sámi, Scottish Gaelic, Irish (Irish Gaelic), Swedish, Danish, Norwegian, Mandarin (and other Chinese languages), and Japanese.

## References

- Baekeland, F., & Lundwall, L. (1975). Dropping out of treatment: A critical review. *Psychological Bulletin*, *82*, 738–783. doi:[10.1037/h0077132](https://doi.org/10.1037/h0077132)
- Baldwin, S. A., Berkeljon, A., Atkins, D. C., Olsen, J. A., & Nielsen, S. L. (2009). Rates of change in naturalistic psychotherapy: Contrasting dose-effect and good-enough level models of change. *Journal of Consulting and Clinical Psychology*, *77*, 203–211. doi:[10.1037/a0015235](https://doi.org/10.1037/a0015235)
- Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: Findings and methods. In M. J. Lambert (Ed.), *Bergin & Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 258–297). Hoboken, NJ: Wiley.
- Barkham, M., Connell, J., Stiles, W. B., Miles, J. N. V., Margison, J., Evans, C., & Mellor-Clark, J. (2006). Dose-effect relations and responsive regulation of treatment duration: The good enough level. *Journal of Consulting and Clinical Psychology*, *74*, 160–167. doi:[10.1037/0022-006X.74.1.160](https://doi.org/10.1037/0022-006X.74.1.160)
- Barkham, M., Hardy, G. E., & Mellor-Clark, J. (2010a). *Developing and delivering practice-based evidence: A guide for the psychological therapies*. Oxford: Wiley-Blackwell.
- Barkham, M., & Margison, F. (2007). Practice-based evidence as a complement to evidence-based practice: From dichotomy to chiasmus. In C. Freeman & M. Power (Eds.), *Handbook of evidence-based psychotherapies: A guide for research and practice* (pp. 443–476). Chichester: Wiley.
- Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., ... & McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Towards practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology*, *69*, 184–196. doi:[10.1037/0022-006X.69.2.184](https://doi.org/10.1037/0022-006X.69.2.184)
- Barkham, M., & Mellor-Clark, J. (2003). Bridging evidence-based practice and practice-based evidence: Developing a rigorous and relevant knowledge for the psychological therapies. *Clinical Psychology & Psychotherapy*, *10*, 319–327. doi:[10.1002/cpp.379](https://doi.org/10.1002/cpp.379)
- Barkham, M., Mellor-Clark, J., Connell, J., Evans, R., Evans, C., & Margison, F. (2010b). The CORE measures & CORE system: Measuring, monitoring, and managing quality evaluation in the psychological therapies. In M. Barkham, G. E. Hardy, & J. Mellor-Clark (Eds.), *Developing and delivering practice-based evidence: A guide for the psychological therapies* (pp. 175–219). Chichester: Wiley.
- Barkham, M., Stiles, W. B., Connell, J., & Mellor-Clark, J. (2012). The meaning of treatment effectiveness in routine NHS primary care psychological therapy services. *Psychology and Psychotherapy: Theory, Research and Practice*, *85*, 1–16. doi:[10.1111/j.2044-8341.2011.02019.x](https://doi.org/10.1111/j.2044-8341.2011.02019.x)
- Barkham, M., Stiles, W. B., Lambert, M. J., & Mellor-Clark, J. (2010). Building a rigorous and relevant knowledge-base for the psychological therapies. In M. Barkham, G. E. Hardy, & J. Mellor-Clark (Eds.), *Developing and delivering practice-based evidence: A guide for the psychological therapies* (pp. 21–61). Chichester: Wiley.
- Barrett, M. S., Chua, W., Crits-Christoph, P., Gibbons, M. B., & Thompson, D. (2008). Early withdrawal from mental health treatment: Implications for psychotherapy practice. *Psychotherapy Theory, Research, Practice, Training*, *45*, 247–267. doi:[10.1037/0033-3204.45.2.247](https://doi.org/10.1037/0033-3204.45.2.247)
- Bickman, L. (2008). A measurement feedback system (MFS) is necessary to improve mental health outcomes. *Journal of the American Association of Child and Adolescent Psychiatry*, *47*, 1114–1119. doi:[10.1097/CHI.0b013e3181825af8](https://doi.org/10.1097/CHI.0b013e3181825af8)
- Bickman, L., Kelley, S., Breda, C., Vides de Andrade, A. R., & Riemer, M. (2011). Effects of routine feedback to clinicians on youth mental health outcomes: A randomized cluster design. *Psychiatric Services*, *62*, 1423–1429. doi:[10.1176/appi.ps.002052011](https://doi.org/10.1176/appi.ps.002052011)
- Brown, J., Dreis, S., & Nace, D. K. (1999). What really makes a difference in psychotherapy outcome? Why does managed care want to know? In M. A. Hubble, B. L. Duncan, & S. D. Miller (Eds.), *The heart and soul of change*. Washington DC: American Psychological Association.
- Budge, S., Baardseth, T. P., Wampold, B. E., & Fluckiger, C. (2010). Researcher allegiance and supportive therapy: Pernicious effects on results of randomized clinical trials. *European Journal of Psychotherapy & Counselling*, *12*, 23–39. doi:[10.1080/13642531003637742](https://doi.org/10.1080/13642531003637742)

- Cahill, J., Barkham, M., & Stiles, W. B. (2010). Systematic review of practice-based research on psychological therapies in routine clinic settings. *British Journal of Clinical Psychology, 49*, 421–454. doi:10.1348/014466509X470789
- Castonguay, L. G., Barkham, M., Lutz, W., & McAleavey, A. (2013). Practice-oriented research: Approaches and applications. In M. J. Lambert (Ed.), *Bergin & Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 85–133). Hoboken, NJ: Wiley.
- Chiesa, M., Wright, M., & Neeld, R. (2003). A description of an audit cycle of early dropouts from an inpatient psychotherapy unit. *Psychoanalytic Psychotherapy, 17*, 138–149. doi:10.1080/1474973031000105294
- Clarkin, J. F., Levy, K. N., Lenzenweger, M. F., & Kernberg, O. F. (2004). The Personality Disorder Institute/Borderline Personality Disorder Research Foundation randomized control trial for borderline personality disorder: Rationale, methods, and patient characteristics. *Journal of Personality Disorders, 18*, 52–72. doi:10.1521/pedi.18.1.52.32769
- Crawford, M. J., Robotham, D., Thana, L., Patterson, S., Weaver, T., Barber, R., Wykes, T., & Roze, D. (2011). Selecting outcome measures in mental health: The views of service users. *Journal of Mental Health, 20*, 336–346. doi:10.3109/09638237.2011.577114
- Crits-Christoph, P., Baranckie, K., Kurcias, J., Beck, A., Carroll, K., Perry, K., ... & Zitrin, C. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research, 1*, 81–91. doi:10.1080/10503309112331335511
- Elfström, M. L., Evans, C., Lundgren, J., Johansson, B., Hakeberg, M., & Carlsson, S. G. (2013). Validation of the Swedish version of the Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM). *Clinical Psychology & Psychotherapy, 20*, 447–455.
- Elkin, I., Falconnier, L., Martinovitch, Z., & Mahoney, C. (2006). Therapist effects in the National Institute of Mental Health Treatment of Depression Collaborative Research program. *Psychotherapy Research, 16*, 144–160. doi:10.1080/10503300500268540
- EuroQol Group. (1990). EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy, 16*, 199–208. doi:10.1016/0168-8510(90)90421-9
- Evans, C., Connell, J., Barkham, M., Margison, F., Mellor-Clark, J., McGrath, G., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry, 180*, 51–60. doi:10.1192/bjp.180.1.51
- Evans, C., Connell, J., Barkham, M., Marshall, C., & Mellor-Clark, J. (2003). Practice-based evidence: Benchmarking NHS primary care counselling services at national and local levels. *Clinical Psychology & Psychotherapy, 10*, 374–388. doi:10.1002/cpp.384
- Falkenström, F., Granström, F., & Holmqvist, R. (2013a). Therapeutic alliance predicts symptomatic improvement session by session. *Journal of Counseling Psychology, 60*, 317–328. doi:10.1037/a0032258
- Falkenström, F., Granström, F., Holmqvist, R., & Psychotherapy Research. (2013b). Working alliance predicts psychotherapy outcome even while controlling for prior symptom improvement. *Psychotherapy Research*. Online publication. doi:10.1080/10503307.2013.847985
- Gilbody, S., Richards, D. A., & Barkham, M. (2007). Diagnosing depression in primary care using self-completed instruments: a UK validation of the PHQ-9 and CORE-OM. *British Journal of General Practice, 57*, 650–652.
- Hatcher, R. L., & Gillaspay, J. A. (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research, 16*, 12–25. doi:10.1080/10503300500352500
- Hellerstein, D. J., & Markowitz, J. C. (2008). Developing supportive psychotherapy as evidence-based treatment. *American Journal of Psychiatry, 165*, 1355–1356. doi:10.1176/appi.ajp.2008.08040565
- Hoffman, J. J. (1985). Client factors related to premature termination of psychotherapy. *Psychotherapy: Theory, Research, Practice, Training, 22*, 83–85. doi:10.1037/h0088530
- Holmqvist, R., Philips, B., & Mellor-Clark, J. (2013). Client and therapist congruence about the client's problems, treatment alliance, and outcome. Manuscript.
- Holmqvist, R., Ström, T., & Foldemo, A. (2013). The effects of psychological treatment in primary care. *Nordic Journal of Psychiatry, 64*, 1–9.
- Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist, 41*, 159–164. doi:10.1037/0003-066X.41.2.159
- Hunsley, J., & Lee, C. M. (2007). Research-informed benchmarks for psychological treatments. Efficacy studies, effectiveness studies, and beyond. *Professional Psychology: Research and Practice, 38*, 21–33. doi:10.1037/0735-7028.38.1.21
- Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist, 63*, 146–159. doi:10.1037/0003-066X.63.3.146
- Kazdin, A. E., & Mazurick, J. L. (1994). Dropping out of child psychotherapy: Distinguishing early and late dropouts over the course of treatment. *Journal of Consulting and Clinical Psychology, 62*, 1069–1075. doi:10.1037/0022-006X.62.5.1069
- Kim, D.-M., Wampold, B. E., & Bolt, D. M. (2006). Therapist effects in psychotherapy: A random-effects modeling of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data. *Psychotherapy Research, 16*, 161–172. doi:10.1080/10503300500264911
- Klein, E. B., Stone, W. N., Hicks, M. W., & Pritchard, I. L. (2003). Understanding dropouts. *Journal of Mental Health Counseling, 25*, 89–100.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*, 606–613. doi:10.1046/j.1525-1497.2001.016009606.x
- Lambert, M. J. (2010). *Prevention of treatment failure: The use of measuring, monitoring, and feedback in clinical practice*. Washington DC: American Psychologist Association.
- Larsson, M., Andersson, G., & Holmqvist, R. (2013). Rupture-repair events in psychological treatments in primary care. Manuscript.
- Larsson, B. P. M., Kaldö, V., & Broberg, A. G. (2009). Similarities and differences between practitioners of psychotherapy in Sweden: A comparison of attitudes between psychodynamic, cognitive, cognitive-behavioral, and integrative therapists. *Journal of Psychotherapy Integration, 19*, 43–66.
- Larsson, M., Andersson, G., & Holmqvist, R. (2013). Rupture-repair events in psychological treatments in primary care. Manuscript.
- Leach, C., Lucock, M., Barkham, M., Stiles, W. B., Noble, R., & Iveson, S. (2006). Transforming between Beck Depression Inventory and CORE-OM scores in routine clinical practice. *British Journal of Clinical Psychology, 45*, 153–166. doi:10.1348/014466505X35335
- Luborsky, L., McLellan, A. T., Diguier, L., Woody, G., & Seligman, D. A. (1997). The psychotherapist matters: Comparison of outcomes across twenty-two therapists and

- seven patient samples. *Clinical Psychology: Science and Practice*, 4, 53–65. doi:10.1111/j.1468-2850.1997.tb00099.x
- Lutz, W., Leach, C., Barkham, M., Lucock, M., Stiles, W. B., Evans, C., ... & Iveson, S. (2005). Predicting rate and shape of change for individual clients receiving psychological therapy: Using growth curve modeling and nearest neighbor technologies. *Journal of Consulting and Clinical Psychology*, 73, 904–913. doi:10.1037/0022-006X.73.5.904
- Mavranzeouli, I., Brazier, J., Rowan, D., & Barkham, M. (2013). Estimating a preference-based index from the Clinical Outcomes in Routine Evaluation—Outcome Measure (CORE-OM): valuation of CORE-6D. *Medical Decision Making*, 33, 381–395. doi:10.1177/0272989X12464431
- Mavranzeouli, I., Brazier, J., Young, T., & Barkham, M. (2011). Using Rasch analysis to form plausible health states amenable to valuation: the development of CORE-6D and CORE-Utility from CORE-OM in order to elicit preferences for common mental health problems. *Quality of Life Research*, 20, 321–333. doi:10.1007/s11136-010-9768-4
- McHugh, R. K., & Barlow, D. H. (2012). *Dissemination and implementation of evidence-based psychological interventions*. New York: Oxford University Press.
- McMuran, M., Huband, N., & Overton, E. (2010). Non-completion of personality disorder treatments: A systematic review of correlates, consequences, and interventions. *Clinical Psychology Review*, 30, 277–287. doi:10.1016/j.cpr.2009.12.002
- Mellor-Clark, J., & Barkham, M. (2012). Using the CORE System to support service development. In C. Feltham & I. Horton (Eds.), *Handbook of counselling and psychotherapy* (3rd ed.). London: Sage.
- Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. H. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *British Journal of Psychiatry*, 180, 461–464. doi:10.1192/bjp.180.5.461
- Novalis, P. N., Rojcewicz, S. J., & Peele, R. (1993). *Clinical manual of supportive psychotherapy*. Washington DC: American Psychiatric Publishing.
- O'Brien, A., Fahmy, R., & Singh, S. P. (2009). Disengagement from mental health services: A literature review. *Journal of Social Psychiatry and Psychiatric Epidemiology*, 44, 558–568. doi:10.1007/s00127-008-0476-0
- O'Donovan, A., Bain, J. D., & Dyck, M. J. (2005). Does clinical psychology education enhance the clinical competence of practitioners? *Professional Psychology: Research and Practice*, 36, 104–111. doi:10.1037/0735-7028.36.1.104
- Philips, B. (2009). Comparing apples and oranges: How do patient characteristics and treatment goals vary between different forms of psychotherapy? *Psychology and Psychotherapy*, 82, 323–336. doi:10.1348/147608309X431491
- Pingitore, D. P., Scheffler, R. M., Schwalm, D., Zarin, D. A., & West, J. C. (2002). Variation in routine psychiatric workload: The role of financing source, managed care participation, and mental health workforce competition. *Mental Health Services Research*, 4, 141–150. doi:10.1023/A:1019759029241
- Pinsker, H. A. (1997). *Primer of supportive therapy*. Hillside, NJ: Analytic Press.
- Reis, B. F., & Brown, L. G. (2006). Preventing therapy dropout in the real world: The clinical utility of videotape preparation and client estimate of treatment duration. *Professional Psychology: Research and Practice*, 37, 311–316. doi:10.1037/0735-7028.37.3.311
- Royal College of Psychiatrists. (2011). *National Audit of Psychological Therapies for Anxiety and Depression*. National Report.
- Saxon, D., & Barkham, M. (2012). Patterns of therapist variability: Therapist effects and the contribution of patient severity and risk. *Journal of Consulting and Clinical Psychology*, 80, 535–546. doi:10.1037/a0028898
- Sinclair, M., & Holmqvist, R. (2013). Congruence in the therapists' and clients' ratings of the therapeutic alliance does not predict outcome in psychological treatment. Manuscript.
- Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, 22, 1092–1097. doi:10.1001/archinte.166.10.1092
- Stiles, W. B., Barkham, M., Connell, J., & Mellor-Clark, J. (2008). Responsive regulation of treatment duration in routine practice in United Kingdom primary care settings: Replication in a larger sample. *Journal of Consulting and Clinical Psychology*, 76, 298–305. doi:10.1037/0022-006X.76.2.298
- Stiles, W. B., Barkham, M., Mellor-Clark, J., & Connell, J. (2008). Effectiveness of cognitive-behavioural, person-centred, and psychodynamic therapies in UK primary care routine practice: Replication in a larger sample. *Psychological Medicine*, 38, 677–688.
- Stiles, W. B., Barkham, M., Twigg, E., Mellor-Clark, J., & Cooper, M. (2006). Effectiveness of cognitive-behavioural, person-centred, and psychodynamic therapies as practiced in UK National Health Service settings. *Psychological Medicine*, 36, 555–566. doi:10.1017/S0033291706007136
- Stiles, W. B., Leach, C., Barkham, M., Lucock, M., Iveson, S., Shapiro, D. A., Iveson, M., & Hardy, G. E. (2003). Early sudden gains in psychotherapy under routine clinic conditions: Practice-based evidence. *Journal of Consulting and Clinical Psychology*, 71, 14–21. doi:10.1037/0022-006X.71.1.14
- Stirman, S. W., DeRubeis, R. J., Crits-Christoph, P., & Brody, P. E. (2003). Are samples in randomized controlled trials of psychotherapy representative of community outpatients? A new methodology and initial findings. *Journal of Consulting and Clinical Psychology*, 71, 963–972. doi:10.1037/0022-006X.71.6.963
- Street, L. L., Niederehe, G., & Lebowitz, B. D. (2000). Toward greater public health relevance for psychotherapeutic intervention research: An NIMH workshop report. *Clinical Psychology: Science and Practice*, 7, 127–137. doi:10.1093/clipsy.7.2.127
- Strupp, H. H., Horowitz, L., & Lambert, M. J. (Eds.). (1997). *Measuring patient changes in mood, anxiety, and personality disorders: Toward a core battery*. Washington DC: American Psychological Association
- Stulz, N., Lutz, W., Leach, C., Lucock, M., & Barkham, M. (2007). Shapes of early change in psychotherapy under routine outpatient conditions. *Journal of Consulting and Clinical Psychology*, 75, 864–874. doi:10.1037/0022-006X.75.6.864
- Tanielian, T. L., Marcus, S. C., Suarez, A. P., & Pincus, H. A. (2001). Trends in psychiatric practice. 1988–1998: II. Case-load and treatment characteristics. *Psychiatric Services*, 52, 880. doi:10.1176/appi.ps.52.7.880
- Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., ... & Stewart-Brown, S. (2009). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes*, 5, 63. doi:10.1186/1477-7525-5-63
- Waller, G. (2009). Evidenced-based treatment and therapist drift. *Behaviour Research and Therapy*, 47, 119–127. doi:10.1016/j.brat.2008.10.018
- Wampold, B. E., & Brown, G. S. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology*, 73, 914–923. doi:10.1037/0022-006X.73.5.914

- Waskow, I. E., & Parloff, M. B. (1975). *Psychotherapy change measures*. Washington DC: US Government Printing Office.
- Wierzbicki, M., & Pekarik, G. (1993). A meta-analysis of psychotherapy dropout. *Professional Psychology: Research and Practice*, 24, 190–195. doi:[10.1037/0735-7028.24.2.190](https://doi.org/10.1037/0735-7028.24.2.190)

- Winston, A., Rosenthal, R., & Pinsker, H. (2012). *Learning supportive therapy. An illustrated guide*. Arlington, VA: American Psychiatric Association.
- Young, A. S., Grusky, O., Jordan, D., & Belin, T. R. (2000). Routine outcome monitoring in a public mental health system: The impact of patients who leave care. *Psychiatric Services*, 51, 85–91.