A CORE Approach to Progress Monitoring and Feedback: Enhancing Evidence and Improving Practice

Michael Barkham
University of Sheffield

John Mellor-Clark
CORE Information Management Systems Ltd., Rugby, United Kingdom

William B Stiles
Miami University and Appalachian State University

This article describes the Clinical Outcomes in Routine Evaluation (CORE) System and reports on its scientific yield and practice impact. First, we describe the suite of CORE measures, including the centerpiece CORE-Outcome Measure (CORE-OM), its short forms, special purpose forms, translations, and psychometric properties, along with the pretreatment CORE Therapy Assessment Form and the CORE End of Therapy Form. Second, we provide an overview of the scientific yield arising from analyses of large CORE data sets collected in routine practice. Third, we describe the use of CORE measures for feedback in practice settings. Finally, we consider future directions for monitoring and feedback in research and practice.

Keywords: CORE-OM, CORE System, patient monitoring, feedback, practice-based evidence

The advent of technologies for routine monitoring of client response to treatment, rapid analysis, and feedback of results raises the questions of what to monitor, how to aggregate the data, and who should receive feedback. The Clinical Outcomes in Routine Evaluation (CORE) System is intended for monitoring psychological change, aggregating on multiple levels, and providing feedback to a variety of stakeholders (Barkham et al., 2001; Evans et al., 2000; Mellor-Clark, Barkham, Connell, & Evans, 1999). Progress can be monitored at the level of treatment services, therapists, clients, sessions, and episodes within sessions. Feedback may be usefully delivered to clients, therapists, managers, service designers, and policymakers, as well as to the scientists who seek to understand psychological change and the clinical theorists who design the treatments.

In this article we offer a summary of monitoring and feedback to both scientific and practice communities arising from the CORE System. We refer to this work as *practice-based evidence*

(Barkham, Hardy, & Mellor-Clark, 2010; Margison et al., 2000), a paradigm akin to practice-oriented research (cf. Castonguay, Barkham, Lutz, & McAleavy, 2013). First, we provide an outline of the CORE measures and system for monitoring progress. Second, we document the yield for scientists and clinical theorists interested in psychological therapies. Third, we highlight feedback procedures for administrators interested in the delivery of services and for participants in therapy. Finally, we offer suggestions regarding future directions and recommendations for the field.

Monitoring: CORE Instruments and System

A team of scientist-practitioners in the United Kingdom developed the CORE Outcome Measure (CORE-OM) and associated CORE System (Barkham et al., 2001; Evans et al., 2000; Mellor-Clark et al., 1999). In this section we provide a brief overview of the suite of CORE measures (for a full account, see Barkham, Mellor-Clark, et al. (2010). All CORE measures are copyleft, and as of 2015, were released under the Creative Commons Attribution-No Derivatives 4.0 International License (see https://www.coresystemtrust.org.uk). This means that the measures are free to use in all formats, including incorporation into electronic systems, providing they are not changed. The CORE group has a commitment to facilitating the widespread adoption of the measures and developing translations. The aim is to reduce barriers and facilitate simple comparability (Waskow, 1975). In addition, there is a commitment to continuing development and innovation.

CORE-OM

The original and central component of the CORE System is the CORE-OM, a 34-item pantheoretical measure of psychological distress and change tapping the domains of Subjective Well-Being (four items), Problems (12 items), Life Functioning (12 items), and

Michael Barkham, Centre for Psychological Services Research, University of Sheffield; John Mellor-Clark, CORE Information Management Systems Ltd., Rugby, United Kingdom; William B Stiles, Miami University and Appalachian State University.

Michael Barkham was a member of the research teams that developed the CORE-OM and CORE-10 measures. He is a CORE System Trustee but receives no financial return from the implementation of the CORE measures or System.

John Mellor-Clark is founder and managing Director of CORE IMS Ltd. As an organization, CORE IMS offers a range of fee-based support services for users of the CORE System (and other measurement tools) that includes: training, software, and consultancy.

Correspondence concerning this article should be addressed to Michael Barkham, Centre for Psychological Services Research, Department of Psychology, Western Bank, University of Sheffield S10 2TN UK. E-mail: m.barkham@sheffield.ac.uk

Risk (six items: four risk-to-self items and two risk-to-others items; Evans et al., 2002). Particular features of the CORE-OM include a balance of high-intensity (18) and low-intensity (16) items to capture a broader intensity of affect and impact; positively keyed items (eight) as well as negatively keyed items (26) as a hedge against acquiescence bias; and items focusing on risk to self and to others. These domains were included for content validity, to represent aspects of clients' experiences deemed central by practitioners for evaluating psychological change in therapy; they are not independent factors. Domain scores other than for Risk intercorrelate strongly, whereas scores on the Risk domain are much less correlated with scores in the other domains (Evans et al., 2002). Although the CORE-OM was developed from the views of practitioners rather than service users, it represents the views of users as well as or better than many other standard client outcome measures (Crawford et al., 2011).

Psychometrics

Reliability. In a general population sample, the internal consistency coefficient alpha was 0.91 ($N = 535$; Connell, Barkham, Stiles, et al., 2007), and in primary care and secondary care samples, it was 0.93 ($N = 5,733$) and 0.95 ($N = 1,918$), respectively (Barkham, Gilbert, Connell, Marshall, & Twigg, 2005). The internal consistency of the domains has been reported as follows: Well-Being = 0.70, Problems = 0.87, Functioning = 0.85, and Risk = 0.77 (Evans et al., 2002). In a clinical waiting-list population ($N = 1,684$) the test–retest reliability was .88 ($n = 119$) at 1 month, .81 ($n = 283$) at 2 months, .83 ($n = 287$) at 3 months, and .80 ($n = 281$) at 4 months (Barkham, Mullin, Leach, Stiles, & Lucock, 2007).

Convergent validity. The CORE-OM correlates highly with versions of the Beck Depression Inventory (BDI): the first version of the BDI, 0.85 to 0.86; BDI-II, 0.75 to 0.81 (Cahill et al., 2006; Evans et al., 2002; Leach et al., 2006). Transformation tables have been produced to convert between the first version of the BDI and CORE-OM scores (Leach et al., 2006). The CORE-OM and Patient Health Questionnaire-9 (PHQ-9; Kroenke, Spitzer, & Williams, 2001) have been evaluated against the Structured Clinical Interview for DSM (SCID; Spitzer, Williams, Gibbon, & First, 1992) diagnosis of depression, resulting in the area under the curve for the CORE-OM of 0.92, and a depression cutoff score of 13, compared with 0.94 for the PHQ-9 (Gilbody, Richards, & Barkham, 2007). This illustrates that depression, whether interviewer-rated or self-rated, correlates very strongly with the more general self-rating of distress and dysfunction assessed by the CORE-OM mean item score.

Scoring and Cutoff Scores

CORE-OM forms are considered valid if three or fewer items are omitted. CORE-OM clinical scores are computed as the mean of completed items multiplied by 10, so clinically meaningful differences are represented by whole numbers. Thus, scores can range from 0 to 40. The CORE-OM's recommended *clinical cutoff score* of 10 was selected to discriminate optimally between a clinical sample and a systematic general population sample (Connell, Barkham, Stiles, et al., 2007). The reliable change index (threshold for change not attributable to chance at $p < .05$) approximates to 5 points, although it varies slightly across samples. Table 1 presents details of statistics from studies contributing to the calculation of these indices.

Alternative CORE-OM Forms

Several CORE short forms, versions for special populations, and translation into languages other than English have been developed, drawing on the original CORE-OM items.

Short versions. In the original development work on the CORE-OM, two parallel 18-item CORE Short Forms A and B were derived from the CORE-OM for alternate repeated administrations in therapy sessions with the aim of reducing the effect of response bias arising from the repeated use of identical items (Barkham, Mellor-Clark, et al., 2010; Evans et al., 2002). Each comprised four subjective well-being items common to both forms and 14 other items, all drawn from the CORE-OM, and differing in each version.

A still briefer version, the CORE-10 (10 items), sampled items from symptoms (weighted toward depression and anxiety), life

Table 1
Psychometrics and Sources Relating to Reliable and Clinically Significant Change

Reliable & clinical change indices	Values	Sample descriptors	Source
Internal consistency (coefficient alpha)	.93 (Primary care)	$N = 5,733$; $M = 18.1$ ($SD = 6.7$)	Barkham et al. (2005)
	.95 (Secondary care)	$N = 1,918$; $M = 1.81$ ($SD = 7.4$)	
Test–retest reliability	.88 (1 month)	$N = 119$; Month 1, $M = 19.36$ ($SD = 7.65$); Month 2, $M = 19.26$ ($SD = 8.24$)	Barkham et al. (2007)
	.81 (2 months)	$N = 283$; Month 1, $M = 20.64$ ($SD = 7.48$); Month 3, $M = 19.53$ ($SD = 7.79$)	
Reliable change index	5.9 (Clinical sample)	$N = 10,761$; $M = 18.3$ ($SD = 7.1$)	Connell, Barkham, Stiles, et al. (2007)
	4.9	$N = 5,613$; Pretherapy, $M = 17.60$ ($SD = 6.33$); Posttherapy, $M = 8.77$ ($SD = 6.43$)	Stiles, Barkham, Mellor-Clark, et al. (2008)
	4.5	$N = 12,746$; Pretherapy, $M = 18.8$ ($SD = 5.1$); Posttherapy, $M = 8.8$ ($SD = 6.1$)	Stiles, Barkham, Connell, et al. (2008)
Clinical cutoff score	10	Clinical sample, $N = 10,761$, $M = 18.3$ ($SD = 7.1$); General population sample, $N = 535$, $M = 4.8$ ($SD = 4.3$)	Connell, Barkham, Stiles, et al. (2007)
Severity bands	Nonclinical: 0–9	Mild: 10–14 Moderate: 15–19 Moderate/Severe: 20–24	Severe: ≥25

functioning, and risk (Barkham et al., 2013). The 10-item format was designed for initial screening and repeated administration and ease of scoring. For the CORE-10, the clinical cutoff and reliable change index values are 11 and 6, respectively. A five-item version (CORE-5) has been used as a thermometer of psychological distress, with the minimal items making it feasible to be administered as part of a therapeutic conversation (Barkham, Mellor-Clark, et al., 2010).

Population-specific versions. Versions of the CORE-OM have been developed for young people aged 11 to 16 years (YP-CORE; Twigg et al., 2009), for people with learning disabilities (LD-CORE, Brooks & Davies, 2008), and for the general population (GP-CORE; Sinclair, Barkham, Evans, Connell, & Audin, 2005). In addition, six items from the CORE-OM have been identified using Rasch analysis to generate a health utility index for use in cost-effectiveness analyses (CORE-6D; Mavranezouli, Brazier, Rowen, & Barkham, 2013; Mavranezouli, Brazier, Young, & Barkham, 2011).

Translations. A major commitment to providing translations for the CORE-OM has so far yielded versions in approximately 25 languages, including 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), Finnish, Mandarin, (and other Chinese languages), and Japanese.[1] A version has also been developed in British Sign Language (Rogers, Evans, Campbell, Young, & Lovell, 2014). For details and availability, see the website https://www.coresystemtrust.org.uk.

CORE Assessment and End of Therapy Forms

The CORE System includes practitioner-completed forms designed to be administered at pre- and posttherapy, termed the CORE Therapy Assessment Form and the CORE End of Therapy Form, respectively (Mellor-Clark & Barkham, 2006). The CORE Therapy Assessment Form comprises referral information, client demographics, and data on the nature, severity, and duration of presenting problems using 14 categories: depression, anxiety, psychosis, personality problems, cognitive/learning difficulties, eating disorder, physical problems, addictions, trauma/abuse, bereavement, self-esteem, interpersonal problems, living/welfare, and work/academic. On the CORE End of Therapy Form, therapists report information about the completed treatment, including number of sessions the client attended, whether the ending was planned or unplanned, and which type(s) of therapy was/were used.

Administration Formats

When launched in 1998, the CORE measures were available only in paper format designed to support optical scanning. However, formats for administration and collation of CORE System measures and clinical services have evolved in step with the technologies increasingly used in routine clinical practice. In 2001, bespoke software was developed for use on personal computers, termed CORE-PC (Version 2.79; CORE Information Management Systems Ltd, 2001). In 2007, this was superseded by cloud-based services under the name CORE Net (Version April 2015; CORE Information Management Systems Ltd, 2006).

The CORE Organization, Network, and Procedures

Following the launch of the CORE System, a support organization—CORE Information Management Systems (CORE IMS; see http://www.coreims.co.uk)—was created to help users of the CORE System maximize the clinical and empirical yield from their data, and the CORE System Trust (see https://coresystemtrust.org.uk) was established to protect ownership of the measures. Early support services focused on data collection procedures and interpretive reports; electronic data collation systems followed later. Users have been invited to donate their CORE data to CORE IMS in accordance with local governance arrangements, and CORE IMS has developed partnerships with investigators to analyze the data for scientific purposes (see next section entitled Generating Knowledge for Stakeholders).

As of 2015, over 25,000 practitioners have been trained in the use of the CORE tools, and the CORE software systems (i.e., CORE PC and CORE Net) have been provided to over 800 organizations. More than 150 services have been supported across nine countries, with an annual caseload of over 150,000 clients treated by a workforce of approximately 5,000 practitioners, for a total of more than 600,000 CORE-measured sessions. Since the release of CORE measures under Creative Commons License in 2015, several major U.K. and European electronic system providers have opted to add CORE to the measurement suites they offer as fee-free measures.

Feedback: Generating Knowledge for Stakeholders

Feedback for Theorists and Treatment Designers: The Scientific Yield

The use of the CORE System for widespread monitoring of psychotherapy and counseling in the United Kingdom has permitted the construction of large, anonymous data sets of clients who have been assessed with components of the CORE System. Between 2001 and 2011, four successive data sets of increasing size were collated. These included 6,610 clients in 2001, 33,587 clients in 2005, 70,245 clients in 2008, and 104,474 clients in 2011, though in each data set, only about one third of the clients had complete data, including both pre- and posttreatment data from the CORE-OM (and short forms), CORE Therapy Assessment Forms, and CORE End of Therapy Forms.

These large data sets provided an opportunity to address questions of treatment effectiveness, including an empirical exploration of the meaning of the term *effectiveness* (Barkham, Stiles, Connell, & Mellor-Clark, 2012) and comparisons of the outcomes of alternative therapeutic approaches (Stiles, Barkham, Mellor-Clark, & Connell, 2008; Stiles, Barkham, Twigg, Mellor-Clark, & Cooper, 2006). Perhaps, surprisingly, in the context of the current dominance of evidence supporting the efficacy of cognitive–behavioral approaches, the replicated results in these very large-*N* CORE studies showed very substantial improvement and recovery rates, but no meaningful differences in outcomes of cognitive–behavioral, person-centered, and psychodynamic therapies,

---

[1] The international program of research focusing on translations of CORE measures is led by Chris Evans (see https://coresystemtrust.org.uk).

whether used alone or in combination with some other approach, in routine mental health care in the United Kingdom This result bolsters the Dodo verdict (everybody has won and all must have prizes) and suggests that theorists and service-delivery designers seeking explanations for variability in client outcomes in routine practice should look elsewhere besides the theoretical orientation of the therapist.

Perhaps, even more surprisingly, in the context of discussions of dose-effect curves and how much therapy is enough, analyses of the relation of treatment duration to outcome in these large data sets has repeatedly shown that, on average, clients who received widely different numbers of sessions had similar recovery and improvement rates (Barkham et al., 2006; Stiles, Barkham, Connell, & Mellor-Clark, 2008; Stiles, Barkham, & Wheeler, 2015). This replicated result was interpreted as suggesting that clients improve at different rates and leave therapy when they have reached a good enough level of psychological well-being. It points theoretical attention toward processes of responsive regulation and client agency. For system administrators, it suggests that attempts to prescribe fixed numbers of sessions may overlook varying needs and appropriate self-regulation by the participants.

Other CORE data-set work has yielded one of the largest studies of therapist effects, applying multilevel modeling to a sample of 119 therapists (Saxon & Barkham, 2012). In addition to the finding that some therapists were twice as effective as others, analyses showed that the size of the therapist effect increased as a function of the initial severity of clients' presenting problems. That is, the more severe the presenting problem for a client, the more it

mattered which therapist saw the client. Figure 1 presents a caterpillar plot (with 95% confidence intervals for each therapist) showing those therapists at each end of the effectiveness spectrum whose client outcomes were consistently and reliably better (left) or worse (right) than average.

A separate research strand has comprised a series of three studies derived from a data set comprising the administration of the 18-item Short Form version of the CORE-OM at each session. First, an investigation of the *sudden gains phenomenon*, initially discovered in cognitive therapy, suggested that sudden gains are as common in other therapies as they are in cognitive therapy (Stiles et al., 2003). Second, an investigation of algorithms for predicting psychotherapy outcomes showed that using a nearest neighbor procedure developed for predicting avalanches in Switzerland was superior to other approaches (Lutz et al., 2005); that is, clinically useful information about a client can be extracted from similar cases (near neighbors) in big data sets. And third, an investigation of the shape of early change in therapy showed five distinct patterns of change that were associated with different outcomes and treatment durations as well as with pretherapy variables (Stulz, Lutz, Leach, Lucock, & Barkham, 2007). For example, clients showing the "early improvement" pattern were initially quite highly impaired but improved a great deal during treatment, even though their treatments were comparatively short. By contrast, clients showing the "high impairment" pattern were also highly impaired initially but improved much less, despite comparatively long treatments. Client age predicted initial CORE-OM scores in the former group but not in the latter group.
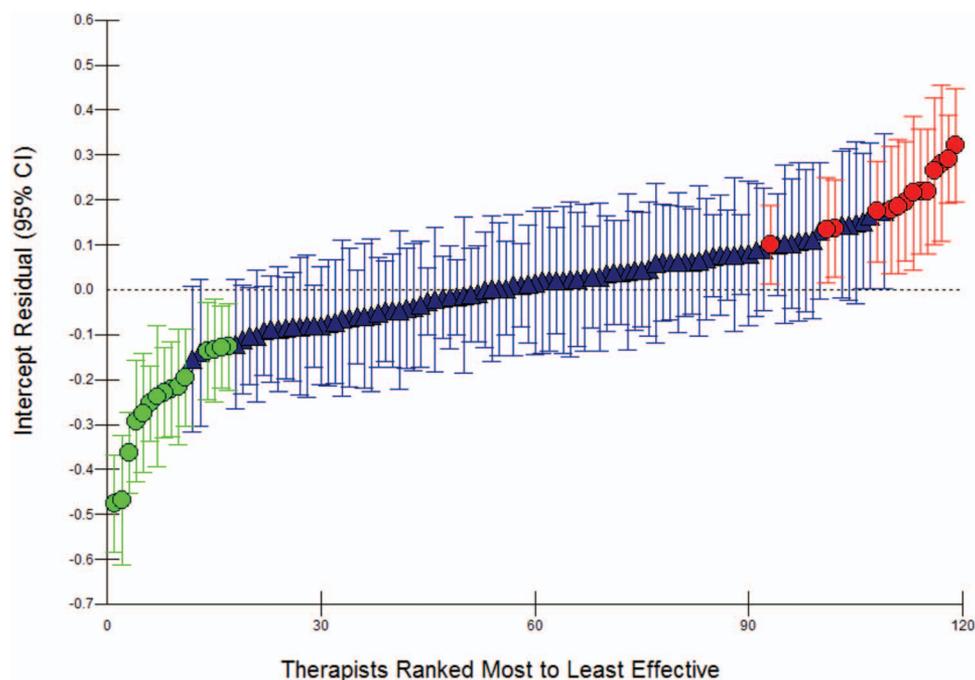

Figure 1. Caterpillar plot of intercept residuals for therapists, ranked, with 95% confidence intervals (CIs). Adapted from "Patterns of Therapist Variability: Therapist Effects and the Contribution of Patient Severity and Risk," by D. Saxon and M. Barkham, 2012, *Journal of Consulting and Clinical Psychology, 80,* p. 540. Copyright 2012 by the American Psychological Association. Adapted with permission. See the online article for the color version of this figure.

Feedback for Health Services Administrators

Benchmarking.   CORE researchers have been active in the development of benchmarking as a way to evaluate change (e.g., Barkham et al., 2001). Benchmarks are reference statistics derived from relevant populations and provide a standard of comparison against which practitioners can assess the quality of their own services. CORE benchmarks have been published for recovery rates (Mullin, Barkham, Mothersole, Bewick, & Kinder, 2006), outcomes in National Health Service (NHS) primary care services (Evans, Connell, Barkham, Marshall, & Mellor-Clark, 2003), and in student services (Connell, Barkham, & Mellor-Clark, 2007). The methodology, indicators, and CORE benchmarks were profiled in a special issue of *Counselling and Psychotherapy Research* (March, 2006). Experiential accounts of how managers subsequently used these CORE indicators for benchmarking their service quality development efforts were profiled in a special issue of *The European Journal of Psychotherapy and Counselling* (June 2006).

Table 2 shows some benchmarks for quality indicators in primary care, higher education, and employee assistance programs that have been used to provide feedback for practitioners and managers (see http://www.coreims.co.uk/Support_User_Benchmarking.html). The median and interquartile range for each of the indicators enables services to locate themselves against these benchmark data. For example, a primary care service might have a 13% rate of client unplanned endings and would therefore place them in the upper quartile, as a lower rate of unplanned endings is more desirable. Their recovery rate might be 58%, which would also place them in the upper quartile, as a higher recovery rate is more desirable.

Feedback implementation model.   Based on experience providing support to users of CORE measures, and drawing on change management theory and developments in implementation science, a feedback implementation model was developed to guide (external) implementation teams as they introduce routine outcome measurement (ROM) into clinical settings (Mellor-Clark, Cross, Macdonald, & Skjulsvik, 2014). It proposes introducing feedback systems in three distinct phases: preparatory, planning, and data management.

The preparatory phase centers on a ROM survey to assess practitioners' attitudes to sessional measurement. The implementation team mandates a response to the ROM survey from every member of the practitioner organization within a specified time frame. The mandate serves to assess the reaction of the organization's management to the implementation, including their ability to secure survey returns from all clinical, supervisory, and managerial staff. The survey data themselves show practitioners' attitudes to sessional measurement and quantify concerns regarding implementation and the use of outcome data. A successful survey will identify internal champions—individuals who actively support sessional measurement and who can be recruited for an active role in training, peer support, and acquiring positive testimonies from practitioners and clients.

In the planning phase, an elected (internal) Implementation Management Group (IMG) is convened by the implementation team to reflect on the ROM survey results. The creation of an IMG helps focus concerns about ROM implementation within the service rather than on external agencies or on the measurement system itself. An effective IMG can preempt many personal and organizational difficulties, and use peers and individuals' supervisors to make the process of adopting sessional measurement more positive.

In the data management phase, individual practitioners receive feedback and are encouraged to reflect on their own experience to promote best practice. The IMG meets regularly with each prac-

Table 2

Selected U.K. National Performance Benchmarks

Selected U.K. national performance benchmarks	Primary care[a]	Higher education[a]	Employment assistance[b]
Clients (*N*)	60,242	28,237	28,476
Services (*N*)	35	18	6
Sample period	2001–2008	2001–2009	2001–2011
CORE-OM completion at pre- and posttherapy			
Median	41%	35%	67%
Interquartile range	34%–49%	26%–51%	64%–74%
Clients above cutoff pretherapy			
Median	88%	90%	87%
Interquartile range	87%–90%	88%–91%	86%–88%
Clients accepted for therapy			
Median	81%	79%	96%
Interquartile range	76%–91%	70%–87%	95%–96%
Clients with unplanned endings			
Median	23%	23%	15%
Interquartile range	15%–31%	20%–31%	14%–17%
Recovered			
Median	50%	56%	52%
Interquartile range	45%–57%	52%–65%	49%–54%
Reliably-improved/not-recovered			
Median	21%	18%	18%
Interquartile range	18%–22%	14%–22%	17%–20%

[a] http://www.coreims.co.uk/Support_User_Benchmarking.html.   [b] From "Benchmarking Key Service Quality Indicators in UK Employee Assistance Programme Counselling: A CORE System Data Profile," by J. Mellor-Clark, E. Twigg, E. Farrell, and A. Kinder, 2012, *Counselling & Psychotherapy Research, 13*, pp. 14–23. http://org/10.1080/14733145.2012.728235.

titioner, usually monthly, to reflect on the latter's individual results and on rates of completion of CORE measures (CORE Therapy Assessment and End of Therapy Forms as well as sessional CORE-OMs) relative to preagreed targets. When practitioners fall short of agreed targets, the IMG consults with supervisors or others to assess and remedy the shortfall. Once completion rates are satisfactory, outcome data interpretation is provided to individual practitioners. This includes electronic alerts for "off-track" cases, mandating a case note for potential discussion in supervision and/or at case management meetings. Promoting such reflective practice as routine appears to quickly allay concerns and replace them with service-wide discussions on working with an appropriate case mix to help secure optimum outcomes. This implementation model has been successfully deployed across clinical teams ranging in size from 30 to 150 practitioners.

Measuring and monitoring clinical outcomes in supervision (MeMOS). In response to a review by Davidson, Perry, and Bell (2015), who argued for drawing evidence from more severe client groups, Davidson et al. (2015) carried out a stepped wedge randomized controlled trial that examined the clinical effectiveness and cost efficiency of continuous monitoring of progress for clients with a wide range and severity of mental and physical problems. In the MeMOS condition, session-by-session feedback on the CORE-10 was given to practitioners and supervisors for discussion in clinical supervision. If a client failed to show improvement, an alert was sent to both clinician and supervisor. In the comparison standard supervision condition, feedback was given only to the therapist. The format of feedback and the alert were identical and utilized an electronic line graph produced from CORE Net that plotted CORE scores (y-axis) for each therapy session (x-axis). Interpreting the score trajectory was made easy by made easy by denoting color-coded severity bands (see Table 1, bottom row) to show progress over the course of therapy. These progress charts were printed and distributed to therapists monthly. Clients received a copy of their chart if they wished.

The results showed no difference in clinical outcomes and no cost differences associated with the MeMOS condition. However, clients in the MeMOS condition required fewer therapy sessions to obtain a similar outcome. If outcome reflects responsive regulation of treatment duration by participants, as suggested earlier (Barkham et al., 2006; Stiles, Barkham, Connell, et al., 2008; Stiles et al., 2015), then the similar mean outcomes are to be expected, and this efficiency may represent success of the intervention.

Barriers and facilitators of outcome monitoring. Lucock et al. (2015) introduced an outcome monitoring system using the CORE-10 in two services within the U.K. National Health Service. Feedback was given to therapists after Session 4. Outcomes of 202 episodes of therapy that included feedback were compared with benchmark data on 136 episodes of therapy for which feedback was not given to therapists. There was no significant difference in the proportions of clients in the different change categories at final outcome for the feedback study compared with the benchmark data. The study identified problems with the extent to which therapists integrated the measures and feedback into the therapy, and focused on organizational barriers to effective feedback, including lack of administrative support, unfamiliarity with information technology, and complexity of the organization. Although therapists discussed the feedback in supervision and, to some extent, with clients—thereby supporting the feasibility and accept-

ability of setting up a routine system in a complex service—the challenges and barriers were apparent.

Individualized Feedback for Participants

Individualized Patient-Progress System. An international practice research network for personalizing health assessment comprising family therapy clinicians and researchers from three European countries developed a pluralistic measurement approach for client feedback combining nomothetic (including the CORE-OM) and idiographic measures for use in family therapy. In 2012, and using CORE Net as the starting point, they incorporated these measures in collaboration with developers into the Individualized Patient-Progress System (IPPS; Sales & Alves, 2012). The IPPS provided data handling and the visualization of clients' progress. Although using it was time consuming, therapists considered the IPPS a "pleasant" and "positive" experience, being a "useful tool to organize data," "provide summaries of preliminary results," and "understand the interaction between the psychological distress of different family members" (Sales, Alves, Evans, Elliott, & On Behalf of Ipha Group, 2014, p. 186).

Clinical use of CORE-OM by trainees. In response to requirements for U.K. clinical psychology doctorate trainees to demonstrate competences in the use of outcome measures, three training programs created a practice research network to encourage the systematic clinical use of outcome measures (Hughes & Latchford, in press). A representation of graphical tracking of outcome scores is shown in Figure 2, in which, initially, the full CORE-OM was administered, then the Short Forms A and B, followed by the CORE-10, and finishing with the full CORE-OM. Early lessons highlighted the need for support to help trainees assimilate the requisite clinical and technical skills. For example, scanning forms for clients' higher scoring items could help focus conversations on areas causing the greatest distress, and sharing tracking graphs with clients could demonstrate their progress in therapy. In-session discussion of the experiences underlying changes in a client's item scores could show how changes sometimes reflect extratherapeutic factors. In addition, concrete measures can help counter getting lost in the hopelessness that clients were experiencing.

Service evaluation of CORE feedback procedure. Via focus groups and semistructured interviews, Unsworth, Cowie, and Green (2012) assessed the response of therapists and clients to having instant visual feedback summarizing their CORE score trajectories at each attended therapy session. In addition, they evaluated the therapists' responses to having such progress or tracking graphs available in supervision. Analyses of interviews with nine therapists and 10 clients at two distinct services yielded six themes: (a) therapists were initially anxious and resistant, (b) therapists adapt "creatively," (c) outcome measures help the client/therapist relationship, (d) clients perceive visual measures as helpful, (e) CORE scores inform supervision, and (f) proper and ongoing training and support of therapists is necessary.

Tracking responses to items in measures (TRIM). Cross, Mellor-Clark, and Macdonald (2014) introduced the TRIM method, which uses the client's individual item scores rather than the overall mean CORE-OM score for individualized feedback. The feedback is structured as a spreadsheet table with the items reordered into their domains (i.e., Well-being, Symptoms, Functioning, Risk) and color coded, so that items having a negative trajectory are indicated by red,
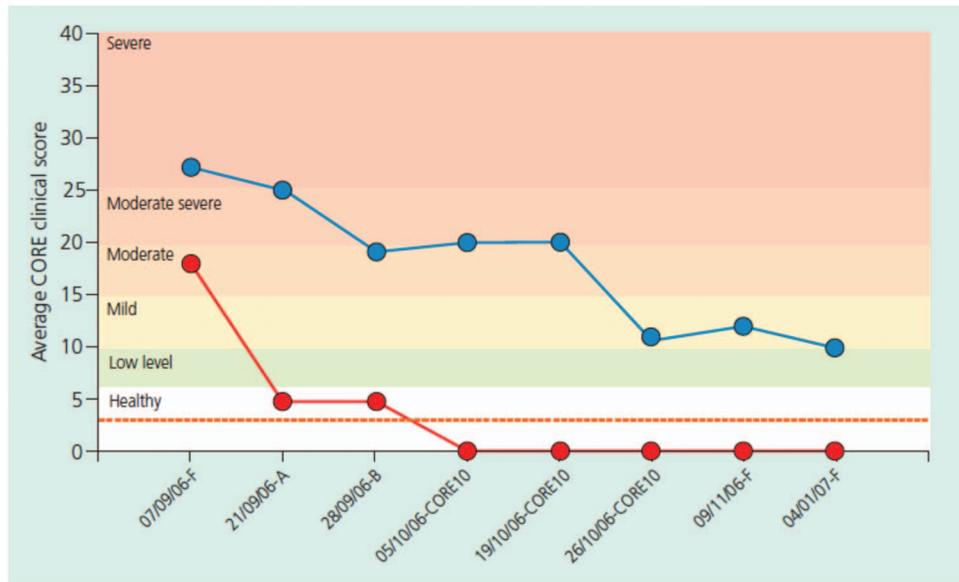
Figure 2. Patient profile using full CORE-OM (F), CORE Short Form A and B, CORE-10. Reprinted from "Managing Therapy Outcomes with CORE Net," by G. Mothersole and T. Jordan, 2007, in P. Gray and J. Mellor-Clark (Eds.), *CORE: A decade of development*, pp. 26–27. Rugby, UK: Penny Gray (publisher). Reprinted with permission. The upper trend line (blue) plots overall psychological distress; the lower trend line (red) plots risk; the dotted line marks the clinical cutoff score for risk. See the online article for the color version of this figure.

and items having a positive trajectory are indicated by green (see Figure 3 for an example). Note that positively worded items are reverse scored. In an analysis of two case studies, Cross et al. (2014) showed that the feedback approach helped motivate the clients to engage more actively in item-change discussions. Tracking and comparing responses to each of the 34 CORE-OM items helped clients identify stressors and triggers that caused them distress, and the focus on response changes to individual items helped foster client hope and maintain morale when only small steps are possible. The approach sometimes highlighted idiosyncratic meanings of item responses, revealing incongruence between what was said in therapy and what was reported in item responses.

The TRIM method meets requirements for the most effective of five categories of procedures for using patient reported outcome measures reviewed by Krägeloh, Czuba, Billington, Kersten, and Siegert (2015). That is, results of the measure are reported back to both practitioner and client, and there is an agreed procedure for discussion of the results that may affect subsequent decisions about the therapy.

Future Work on Progress Monitoring and Feedback

For the measure developer community. Future developments in data collection seem likely to move toward computerized adaptive testing (individualized self-report instruments), drawing on large data banks of items held in the public domain. The combination of such systems with smartphone technology will lead to more efficient and tailored measurement that is responsive to an individual's current and changing psychological state. However, such a strategy needs to adapt the traditional nomothetic approach to item selection with idiographic tools that also have an established tradition within psychological

measurement (see Shapiro, 1961) and have been shown to capture the within treatment variability of individual client change (e.g., Barkham, Stiles, & Shapiro, 1993).

For the research community. So far, the largest CORE data sets include CORE-OM data obtained only at pre- and posttherapy. And because many clients do not complete therapy, endpoint data were obtained on only about one third of clients. With the introduction of sessional measurement, outcomes (calculated as the last obtained session CORE score) are available for significantly higher proportions of cases, particularly including clients who did not complete therapy. Future work with such samples can reexamine comparable effectiveness of different theoretical approaches, equivalent outcomes of clients receiving very different numbers of sessions, and substantial differences in effectiveness of different therapists.

Large-scale monitoring can investigate whether most therapists are generalists, able to work with most clients, or whether each therapist is mainly successful with a relatively small proportion of clients. Careful tracking could also alert practitioners and managers in time to transfer clients across practitioners when lack of progress warrants it.

The trend toward the availability of big data for research purposes permits increasingly sophisticated statistical analyses. For example, multilevel modeling can reflect the nested structure of most therapy data, accounting for the complex case mix, and yielding confidence intervals around each therapist's outcomes that protect inferences from the pitfalls of simplistic performance thresholds.

For the practice community. The feedback implementation model collates experiential data, charting the personal journeys of
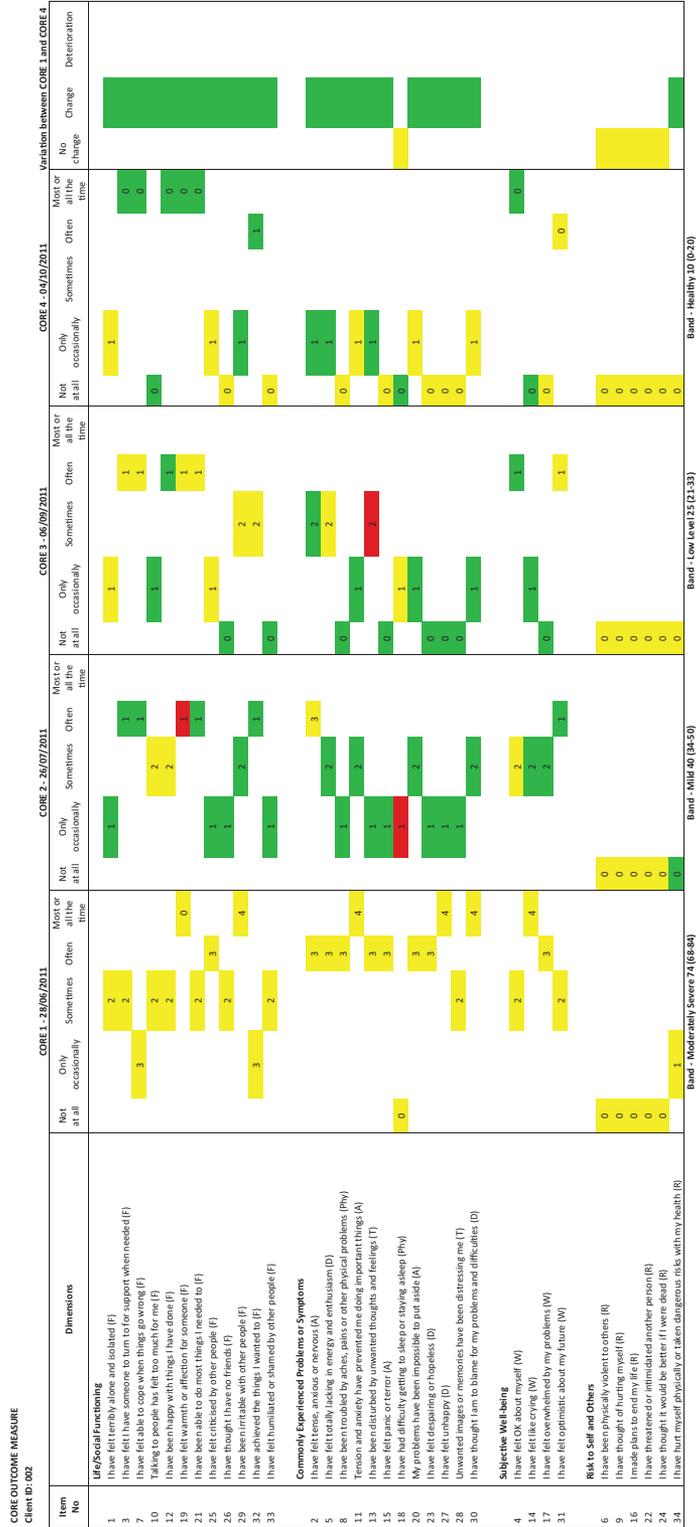
Figure 3. Tracking responses to items in measures chart showing pretreatment, during, and posttreatment scores. Reprinted from "Tracking Responses to Items in Measures as a Means of Increasing Therapeutic Engagement in Clients: A Complementary Clinical Approach to Tracking Outcomes," by S. Cross, J. Mellor-Clark, and J. Macdonald, 2014, *Clinical Psychology and Psychotherapy*, Advance online publication, Wiley Online Library, http://dx.doi.org/10.1002/cpp.1929. Reprinted with permission. See the online article for the color version of this figure.

clinicians. For many practitioners, even (perhaps particularly) for those who are highly experienced, such ROM can be challenging. However, the use of feedback measures needs to move beyond being an administrative activity and rather understood as a clinical skill that can be useful to trainees and experienced therapists alike.

The move toward sessional monitoring and feedback has been a step change in assessing client progress. Our view is that such monitoring needs to be responsive to individual needs and styles. Using identical items at every session with every client may not be the most helpful model. Instead, progress monitoring can be tailored to individual client needs.

An exciting challenge is to blend the microlevel, responsive, moment-by-moment individual feedback to clients with macrolevel analysis of big data. This work offers an opportunity to bridge the scientist-practitioner gap.

References

Barkham, M., Bewick, B. M., Mullin, T., Gilbody, S., Connell, J., Cahill, J., . . . Evans, C. (2013). The CORE-10: A short measure of psychological distress for routine use in the psychological therapies. *Counselling & Psychotherapy Research, 13,* 3–13. http://dx.doi.org/10.1080/14733145.2012.729069

Barkham, M., Connell, J., Stiles, W. B., Miles, J. N. V., Margison, F., 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.

Barkham, M., Gilbert, N., Connell, J., Marshall, C., & Twigg, E. (2005). Suitability and utility of the CORE-OM and CORE-A for assessing severity of presenting problems in psychological therapy services based in primary and secondary care settings. *The British Journal of Psychiatry, 186,* 239–246. http://dx.doi.org/10.1192/bjp.186.3.239

Barkham, M., Hardy, G. E., & Mellor-Clark, J. (Eds.). (2010). *Developing and delivering practice-based evidence: A guide for the psychological therapies*. Chichester, UK: Wiley. http://dx.doi.org/10.1002/9780470687994

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., . . . the Clinical Outcomes in Routine Evaluation-Outcome Measures. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology, 69,* 184–196. http://dx.doi.org/10.1037/0022-006X.69.2.184

Barkham, M., Mellor-Clark, J., Connell, J., Evans, R., Evans, C., & Margison, F. (2010). 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, UK: Wiley. http://dx.doi.org/10.1002/9780470687994.ch8

Barkham, M., Mullin, T., Leach, C., Stiles, W. B., & Lucock, M. (2007). Stability of the CORE-OM and BDI-I prior to therapy: Evidence from routine practice. *Psychology and Psychotherapy: Theory, Research and Practice, 80,* 269–278. http://dx.doi.org/10.1348/147608306X148048

Barkham, M., Stiles, W. B., Connell, J., & Mellor-Clark, J. (2012). Psychological treatment outcomes in routine NHS services: What do we mean by treatment effectiveness? *Psychology and Psychotherapy: Theory, Research and Practice, 85,* 1–16. http://dx.doi.org/10.1111/j.2044-8341.2011.02019.x

Barkham, M., Stiles, W. B., & Shapiro, D. A. (1993). The shape of change in psychotherapy: Longitudinal assessment of personal problems. *Journal of Consulting and Clinical Psychology, 61,* 667–677. http://dx.doi.org/10.1037/0022-006X.61.4.667

Brooks, M., & Davies, S. (2008). Pathways to participatory research in developing a tool to measure feelings. *British Journal of Learning Disabilities, 36,* 128–133. http://dx.doi.org/10.1111/j.1468-3156.2007.00476.x

Cahill, J., Barkham, M., Stiles, W. B., Twigg, E., Hardy, G. E., Rees, A., & Evans, C. (2006). Convergent validity of the CORE measures with measures of depression for clients in brief cognitive therapy for depression. *Journal of Counseling Psychology, 53,* 253–259. http://dx.doi.org/10.1037/0022-0167.53.2.253

Castonguay, L. G., Barkham, M., Lutz, W., & McAleavy, 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.

Connell, J., Barkham, M., & Mellor-Clark, J. (2007). CORE-OM mental health norms of students attending university counselling services benchmarked against an age-matched primary care sample. *British Journal of Guidance & Counselling, 35,* 41–57. http://dx.doi.org/10.1080/03069880601106781

Connell, J., Barkham, M., Stiles, W. B., Twigg, E., Singleton, N., Evans, O., & Miles, J. N. V. (2007). Distribution of CORE-OM scores in a general population, clinical points and comparison with the CIS-R. *The British Journal of Psychiatry, 190,* 69–74. http://dx.doi.org/10.1192/bjp.bp.105.017657

CORE Information Management Systems Ltd. (2001). *CORE-PC (Version 2.79)* [Computer software]. Rugby, UK: Author.

CORE Information Management Systems Ltd. (2006). *CORE Net (Version April 2015)* [Computer software]. Rugby, UK: Author.

Crawford, M. J., Robotham, D., Thana, L., Patterson, S., Weaver, T., Barber, R., . . . Rose, D. (2011). Selecting outcome measures in mental health: The views of service users. *Journal of Mental Health (Abingdon, England), 20,* 336–346. http://dx.doi.org/10.3109/09638237.2011.577114

Cross, S., Mellor-Clark, J., & Macdonald, J. (2014). Tracking responses to items in measures as a means of increasing therapeutic engagement in clients: A complementary clinical approach to tracking outcomes. *Clinical Psychology and Psychotherapy*. Advance online publication. http://dx.doi.org/10.1002/cpp.1929

Davidson, K., Cassells, M., Begley, A., Lloyd, S. M., Barry, S. E., Bell, L., . . . Walker, A. (2015). *A randomised controlled trial assessing the impact of feeding back patient progress in psychological therapy through supervision: The MeMOS project*. Manuscript submitted for publication.

Davidson, K., Perry, A., & Bell, L. (2015). Would continuous feedback of patient's clinical outcomes to practitioners improve NHS psychological therapy services? Critical analysis and assessment of quality of existing studies. *Psychology and Psychotherapy, 88,* 21–37. http://dx.doi.org/10.1111/papt.12032

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *The British Journal of Psychiatry, 180,* 51–60. http://dx.doi.org/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. http://dx.doi.org/10.1002/cpp.384

Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., Audin, K., Connell, J., & McGrath, G. (2000). CORE: Clinical Outcomes in Routine Evaluation. *Journal of Mental Health, 9,* 247–255. http://dx.doi.org/10.1080/713680250

Gilbody, S., Richards, D., & Barkham, M. (2007). Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ-9 and CORE-OM. *The British Journal of General Practice, 57,* 650–652.

Hughes, J., & Latchford, G. (in press). The use of outcome and sessional measures in clinical psychology training. *Clinical Psychology Forum*.

Krägeloh, C. U., Czuba, K. J., Billington, D. R., Kersten, P., & Siegert, R. J. (2015). Using feedback from patient-reported outcome measures in mental health services: A scoping study and typology. *Psychiatric Services, 66,* 224–241. http://dx.doi.org/10.1176/appi.ps.201400141

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. http://dx.doi.org/10.1348/014466505X35335

Lucock, M., Halstead, J., Leach, C., Barkham, M., Tucker, S., Randall, C., . . . Saxon, D. (2015). A mixed method investigation of patient monitoring and feedback in routine practice: Barriers and facilitators. *Psychotherapy Research, 25,* 633–646. http://dx.doi.org/10.1080/10503307.2015.1051163

Lutz, W., Leach, C., Barkham, M., Lucock, M., Stiles, W. B., Evans, C., . . . Iveson, S. (2005). Predicting change for individual psychotherapy clients on the basis of their nearest neighbors. *Journal of Consulting and Clinical Psychology, 73,* 904–913. http://dx.doi.org/10.1037/0022-006X.73.5.904

Margison, F. R., Barkham, M., Evans, C., McGrath, G., Clark, J. M., Audin, K., & Connell, J. (2000). Measurement and psychotherapy. Evidence-based practice and practice-based evidence. *The British Journal of Psychiatry, 177,* 123–130. http://dx.doi.org/10.1192/bjp.177.2.123

Mavranezouli, I., Brazier, J. E., Rowen, 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. http://dx.doi.org/10.1177/0272989X12464431

Mavranezouli, I., Brazier, J., Young, T., & Barkham, M. (2011). Using Rasch analysis to form plausible health states amenable to valuation: The development of CORE-6D from a measure of common mental health problems (CORE-OM). *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 20,* 321–333. http://dx.doi.org/10.1007/s11136-010-9768-4

Mellor-Clark, J., & Barkham, M. (2006). The CORE System: Developing and delivering practice-based evidence through quality evaluation. In C. Feltham & I. Horton (Eds.), *Handbook of counselling and psychotherapy* (2nd ed., pp. 207–224). London, UK: Sage.

Mellor-Clark, J., Barkham, M., Connell, J., & Evans, C. (1999). Practice-based evidence and standardised evaluation: Informing the design of the CORE System. *European Journal of Psychotherapy, Counselling and Health, 3,* 357–374. http://dx.doi.org/10.1080/13642539908400818

Mellor-Clark, J., Cross, S., Macdonald, J., & Skjulsvik, T. (2014). Leading horses to water: Lessons from a decade of helping psychological therapy services use routine outcome measurement to improve practice. *Administration and Policy in Mental Health*. Advance online publication.

Mullin, T., Barkham, M., Mothersole, G., Bewick, B. M., & Kinder, A. (2006). Recovery and improvement benchmarks for counselling and the psychological therapies in routine primary care. *Counselling & Psychotherapy Research, 6,* 68–80. http://dx.doi.org/10.1080/14733140600581515

Rogers, K., Evans, C., Campbell, M., Young, A., & Lovell, K. (2014). The reliability of British Sign Language and English versions of the Clinical Outcomes in Routine Evaluation—Outcome Measure with d/Deaf populations in the UK: An initial study. *Health & Social Care in the Community, 22,* 278–289. http://dx.doi.org/10.1111/hsc.12078

Sales, C. M. D., & Alves, P. C. G. (2012). Individualized patient-progress systems: Why we need to move towards a personalized evaluation of psychological treatments. *Canadian Psychology, 53,* 115–121. http://dx.doi.org/10.1037/a0028053

Sales, C. M. D., Alves, P. C. G., Evans, C., & Elliott, R., & On Behalf of Ipha Group. (2014). The Individualised Patient-Progress System: A decade of international collaborative networking. *Counselling & Psychotherapy Research, 14,* 181–191. http://dx.doi.org/10.1080/14733145.2014.929417

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. http://dx.doi.org/10.1037/a0028898

Shapiro, M. B. (1961). The single case in fundamental clinical psychological research. *British Journal of Medical Psychology, 34,* 255–262. http://dx.doi.org/10.1111/j.2044-8341.1961.tb00949.x

Sinclair, A., Barkham, M., Evans, C., Connell, J., & Audin, K. (2005). Rationale and development of a general population well-being measure: Psychometric status of the GP-CORE in a student sample. *British Journal of Guidance & Counselling, 33,* 153–174. doi:10.1080/03069880500132581

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. http://dx.doi.org/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. http://dx.doi.org/10.1017/S0033291707001511

Stiles, W. B., Barkham, M., Twigg, E., Mellor-Clark, J., & Cooper, M. (2006). Effectiveness of cognitive-behavioural, person-centred and psychodynamic therapies as practised in UK National Health Service settings. *Psychological Medicine, 36,* 555–566. http://dx.doi.org/10.1017/S0033291706007136

Stiles, W. B., Barkham, M., & Wheeler, S. (2015). Duration of psychological therapy: Relation to recovery and improvement rates in UK routine practice. *The British Journal of Psychiatry, 207,* 115–122. http://dx.doi.org/10.1192/bjp.bp.114.145565

Stiles, W. B., Leach, C., Barkham, M., Lucock, M., Iveson, S., Shapiro, D. A., . . . 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. http://dx.doi.org/10.1037/0022-006X.71.1.14

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. http://dx.doi.org/10.1037/0022-006X.75.6.864

Twigg, E., Barkham, M., Bewick, B. M., Mulhern, B., Connell, J., & Cooper, M. (2009). The Young Person's CORE: Development of a brief outcome measure for young people. *Counselling & Psychotherapy Research, 9,* 160–168. http://dx.doi.org/10.1080/14733140902979722

Unsworth, G., Cowie, H., & Green, A. (2012). Therapists' and clients' perceptions of routine outcome measurement in the NHS: A qualitative study. *Counselling & Psychotherapy Research, 12,* 71–80. http://dx.doi.org/10.1080/14733145.2011.565125

Waskow, I. E. (1975). Selection of a core battery. In I. E. Waskow & M. B. Parloff (Eds.), *Psychotherapy change measures* (pp. 245–269). Washington, DC: U.S. Government Printing Office.