

Assessing Therapist Adherence to Recovery-Focused Cognitive Behavioural Therapy for Psychosis Delivered by Telephone with Support from a Self-Help Guide: Psychometric Evaluations of a New Fidelity Scale

Samantha Hartley

University of Manchester and Greater Manchester West Mental Health NHS Foundation Trust, UK

Phillippa Scarratt and Sandra Bucci*

University of Manchester, UK

James Kelly, John Mulligan, Sandra T. Neil and Zoe Rivers

Greater Manchester West Mental Health NHS Foundation Trust, UK

Christopher D. J. Taylor

University of Manchester, and Manchester Mental Health and Social Care Trust, UK

Mary Welford

Greater Manchester West Mental Health NHS Foundation Trust, UK

Gillian Haddock

University of Manchester, and Manchester Mental Health and Social Care Trust, UK

Background: Recent research has highlighted the importance of psychological interventions such as cognitive behavioural therapy (CBT) in improving outcomes and promoting recovery for people with experience of psychosis, although a lack of trained therapists means that

*N.B. Authors 3–9 are listed alphabetically

Reprint requests to Samantha Hartley, Division of Clinical Psychology, 2nd Floor, Zochonis Building, University of Manchester, Brunswick Street, Manchester M13 9PL, UK. E-mail: samantha.hartley@manchester.ac.uk

availability of face-to-face CBT is low. Alternative modes of delivering CBT are being explored, such as telephone and self-help methods, although research to date on whether they can be implemented effectively is limited. **Aims:** The aims of the present study were to describe and evaluate a new therapy fidelity scale (ROSTA; Recovery Oriented Self-help and Telephone therapy Adherence). This scale was developed to assess fidelity to cognitive behaviour therapy for psychosis (CBTp) focused on improving recovery, with optional subscales for delivery over the telephone and alongside a self-help guide. **Method:** Experienced CBT therapists rated recorded therapy sessions using the ROSTA scale. The scores were used to assess internal consistency and inter-rater reliability, before being compared to scores from an independent expert rater using an alternative fidelity scale for cognitive therapy in psychosis (the CTS-Psy), to investigate concurrent validity. **Results:** The ROSTA scale demonstrated excellent internal consistency, inter-rater reliability and validity when evaluated as a whole, although findings were mixed in terms of the individual subscales and items. **Conclusions:** The ROSTA scale is, on the whole, a reliable and valid tool, which may be useful in training and supervision, a utility that would be further emphasized if the therapeutic intervention it assesses is deemed to be efficacious based on future work.

Keywords: Psychosis, recovery, fidelity, cognitive, behavioural, psychometric.

Introduction

Cognitive behavioural therapy for individuals with experience of psychosis (CBTp)

CBTp is now a recommended treatment for all people meeting criteria for a diagnosis of schizophrenia (National Institute for Health and Clinical Excellence [NICE], 2009). CBTp has been shown to be more effective in improving psychotic symptoms than treatment-as-usual and other psychological interventions and demonstrates modest effects on wider outcomes such as functioning and mood (Sarin, Wallin and Widerlov, 2011; Wykes, Steel, Everitt and Tarrier, 2008). NICE currently recommend a minimum of 16 planned sessions per client (NICE, 2009), although a lack of appropriately trained therapists in England and Wales means that the availability of face-to-face CBT is limited (Lovell and Richards, 2000) and demand for CBT currently outweighs supply (Shapiro, Cavanagh and Lomas, 2003), with increasing strain on service resources. The shortage of CBT practitioners presents a challenge both to mental health service providers and policy makers, which has led to the exploration of alternative modes of delivering CBT as a means of facilitating access to therapeutic services. There has been considerable interest in telephone-based CBT in recent years, partly due to the potential for convenience to both service user and therapist, flexibility of location, and its ability to increase therapists' capacity. It has been shown to be a viable and effective way of delivering CBT, displaying similar benefits to CBT delivered face-to-face for problem areas such as obsessive-compulsive disorder (Lovell et al., 2006) and depression (Mohr et al., 2005). The use of self-help materials and guided self-help is also a growing area of interest in working with mental health problems and has been suggested as a possible adjunct to traditional therapeutic sessions, providing a partial solution to the issue of limited availability of accredited therapists described previously (Keeley, Williams and Shapiro, 2002).

Recovery for people with experience of psychosis

Outcomes in CBT for psychosis treatment trials have been varied, although they have tended to focus on symptom removal or symptom reduction (Wykes et al., 2008). However,

research suggests that broader recovery outcomes that reflect the important goals identified by service users may be important to focus on in treatment and may be more meaningful and relevant for service users with psychosis (Pitt, Kilbride, Nothard, Welford and Morrison, 2007). The “recovery movement” (Stickley and Wright, 2011a, b) fosters an idiosyncratic perspective on recovery in mental health care. The concept of recovery in relation to psychosis is a multidimensional process (May, 2004) that values social and psychological recovery, in addition to clinical recovery i.e. “getting rid of symptoms” (Coleman, 1999). Recovery encompasses an on-going, personal process of changing attitudes and feelings toward psychosis, as well as developing new meaning and purpose in life, whilst managing the difficulties associated with symptoms (Anthony, 1993; Pitt et al., 2007).

A recent clinical trial (STAR-T; ISRCTN50487713) has been evaluating a new form of CBTp that is focused on broad recovery outcomes, and is delivered over the telephone with the support of a multi-faceted self-help guide. This approach emphasizes working flexibly with people on their individually defined recovery objectives. This recovery-focused approach to CBTp facilitates the formation of links between components of traditional CBTp (for example, engagement and collaboration) and the establishment of clients’ concept of their own personal recovery goals. The aim of the therapy is therefore to support change that is consistent with the spirit of recovery; in other words, the client’s own concept of this. New outcome measures that focus on these aspects of recovery, such as Subjective Experience of Psychosis Scale (SEPS; Haddock et al., 2011) or Questionnaire about the Process of Recovery (QPR; Neil et al., 2009) are being used within the trial to assess this change.

Treatment fidelity scales for CBTp

With the development of new delivery modes for CBTp, effective fidelity measures are necessary to ensure that therapists can deliver CBTp interventions to the required standard. This will also ensure that the intervention protocol is adhered to (Bellg et al., 2004; Moncher and Prinz, 1991) and facilitate appropriate training and supervision. Treatment fidelity in psychological interventions involves both therapist adherence; the degree to which a therapist utilizes techniques and approaches prescribed by the treatment protocol or manual, and therapist competence; the extent to which the therapist uses their skills to deliver the relevant aspects of the approach whilst responding appropriately to contextual variables (Waltz, Addis, Koerner and Jacobson, 1993). The last 15 years have seen the development of fidelity scales to evaluate CBTp. The Cognitive Therapy Scale for Psychosis (CTS-Psy), developed by Haddock and colleagues (2001), is an example of a treatment fidelity tool designed specifically to evaluate the skills of CBT therapists whilst working with clients with experience of psychosis in individual face-to-face sessions. Additional therapy fidelity scales to assess adherence to CBTp have also been developed and evaluated (for example, Bell et al., 2008; Rollinson et al., 2008; Startup, Jackson and Pearce, 2002), although reports on the psychometric properties of the scales have not always been reported or fully explored, which is essential to promote confidence in their reliability and validity.

A new scale for recovery-focused CBTp delivered by telephone with support from a self-help guide

There are currently no published fidelity scales available to assess therapist adherence to intervention approaches involving self-help materials, with sessions delivered over the

telephone and focused on broad recovery outcomes. Previously validated fidelity measures reflecting different intervention approaches and methods may not be appropriate. As a result, a new treatment fidelity scale, the ROSTA scale (full title: Recovery Oriented CBT for psychosis: supported Self help and Telephone therapy Adherence scale; see Table 1 for a list of items covered), was developed to specifically assess fidelity to this novel delivery mode of recovery-based CBTp. The scale comprises subscales measuring fidelity issues fundamental to recovery focused CBT regardless of modality, for example, interpersonal effectiveness and choice of intervention. It also has modality specific subscales relating to a) the implementation of the therapy alongside a self-help guide and b) as delivered over the telephone. This allows for flexibility in its potential uses for clinicians, supervisors and researchers.

Aims and hypotheses

The aim of the present study was to describe the development of the ROSTA scale, investigate its psychometric properties and examine its utility in evaluating the adherence of therapists engaged in this type of CBTp with people experiencing psychosis. Specifically, we examined the internal consistency and inter-rater reliability of the scale, and its concurrent validity as compared to another similar tool (the CTS-Psy, Haddock et al., 2001). We hypothesized that both the ROSTA scale as a whole and its subscales would display good internal consistency as demonstrated by a Cronbach's alpha coefficient of .6 or above (Landis and Koch, 1977). In addition, we predicted that the scale and each subscale would display good inter-rater reliability when used by expert raters, as determined by a score of .6 or more according to previously established cut-offs for inter-rater reliability (Landis and Koch, 1977). Finally, we hypothesized that the ROSTA scale would demonstrate moderate to good validity against item scores on the CTS-Psy (Haddock et al., 2001), as measured by a significant correlation coefficient ranging from .4 to .8 (Landis and Koch, 1977) between CTS-Psy and ROSTA scale scores. Correlation coefficients that are very high (for example, .9 or more) would indicate that the scales are very similar and would raise questions as to whether the new scale was redundant.

Method

Phase 1: The fidelity scale

Intervention. The intervention for which the therapy fidelity scale was developed consists of the provision of a self-help Recovery Guide, and up to 30 weekly telephone CBT sessions lasting 45 minutes over 9 months, in addition to usual treatment. The Recovery Guide is a self-help resource intended to be supported by therapeutic input delivered by a cognitive-behaviour therapist. The guide was collaboratively developed by people with personal experience of mental health difficulties, researchers, and mental health clinicians, and drew from existing evidence-based cognitive-behavioural interventions for psychosis. Choice regarding which of the many sections of the guide to prioritize is based on collaborative decisions made between client and therapist. The therapy sessions employ cognitive behavioural-interventions for psychosis, while supporting clients to utilize resources from the Recovery Guide. The intervention is a recovery-oriented, person-centred flexible approach that focuses on clients' concepts of recovery and supports them to work towards individual recovery goals using a CBT approach.

Development of the fidelity scale

The ROSTA scale was created as part of an iterative process involving four main stages of development. Research team members involved in this process included clinical psychologists with extensive experience and knowledge of psychosis and cognitive behavioural interventions, and experienced researchers. Prior to developing the new scale, the research team reviewed a number of existing fidelity tools that assess the quality of CBT when working with people experiencing psychosis. It was then established that the scale would be partly based upon the CTS-Psy (Haddock et al., 2001) as this scale is intended for use with cognitive behavioural interventions for people with experience of psychosis, and has been shown to be reliable and valid. The 10 items and 60 sub-items of the CTS-Psy were then examined closely to determine which could be retained as part of the new scale and whether there were any potential items that could be modified to foster a recovery-oriented approach delivered over the telephone with a self-help Recovery Guide. The elements of the new scale were also considered in light of a Delphi study outlining the key aspects of CBTp from clinicians' perspectives (e.g. Morrison and Barratt, 2010) and input from people with lived experience of mental health difficulties. Several components did apply to this new method of delivery as they relate to the core components expected of CBT: session planning, providing and eliciting feedback, understanding, interpersonal effectiveness, collaboration, guided discovery, focus on key cognitive behavioural elements, and choice of intervention and homework. However, a number of the items needed to be added or modified to reflect the recovery ethos of the intervention and the specific mode of delivery.

Three new subscales were added following this process: therapists' skill in using and integrating the guide into the CBTp intervention, therapists' skill at delivering the therapy with the guide over the telephone and facilitating appropriate emotional expression. The last of these new items was added to reflect the skill of acknowledging, drawing out and utilising emotion as part of the therapeutic process, which is an integral part of therapy and has been acknowledged in other fidelity scales (CTS-R; Blackburn et al., 2001). The other items were added to accommodate the two key aspects of this new mode of CBTp, incorporating a self-help guide and delivery of the intervention over the telephone.

A draft version of the ROSTA scale was created and the scale developers collaboratively reviewed and revised the scale in discussion groups comprised of researchers, clinicians and researchers with lived experience of mental health difficulties. This led to a number of alterations, including modifying the wording of various phrases, such as replacing the word "patient" with "client" and rewording "client's recovery" to "client's concept of recovery". These changes reflect the importance of not prescribing the language or concept of recovery but to work within the spirit of recovery. The traditional "homework" element of CBT was replaced with the item "between-session activities" to portray a less instructive and more collaborative approach. A full list of the final subscales and example items contained within each is included in Table 1.

The revised scale was piloted by five clinicians experienced in CBTp, in order to assess the scale's ease of use and appropriateness. Feedback was elicited following these practice ratings, and further minor revisions were made. The redrafted scale was then reviewed by a service user researcher not involved in the initial scale development, who provided suggestions for changes to the wording, focus of items and provided additional insight into the key elements of CBTp from a client perspective.

Table 1. Comparison of the subscales of the CTS-Psy and the ROSTA scale

Cognitive Therapy Scale – Psychosis (CTS-Psy)	Recovery Oriented Self-help and Telephone therapy Adherence (ROSTA) scale (One example item is included beneath each subscale title)
General subscale:	General section:
a. Agenda	1. Session plan: The therapist ensured that each of the agenda items was addressed appropriately and, where possible, links were made between agenda items and the client's recovery goals or understanding of their experiences.
b. Feedback	2. Providing and eliciting feedback: The therapist facilitated the client to discuss and provide feedback in relation to their concept of recovery where appropriate.
c. Understanding	3. Understanding: The therapist acknowledged and placed emphasis on the validity or importance of the client's perspective or experiences.
d. Interpersonal effectiveness	4. Interpersonal effectiveness: The therapist communicated warmth, concern or positive regard during the session where appropriate.
e. Collaboration	5. Collaboration: The therapist asked the client for suggestions on how to proceed and offered choices when feasible in relation to the therapy.
Specific subscale:	Specific section:
f. Guided discovery	6. Guided discovery: The therapist used questions to facilitate the client to generate their own ideas or solutions regarding their personal recovery journey.
g. Focus on key cognitions	7. Focus on key cognitive behavioural elements: The therapist and client collaboratively established the key aspects of the client's thoughts, feelings or behaviours that are most significantly related to the client's current experiences or understanding of recovery.
h. Choice of intervention	8. Choice of intervention: The techniques chosen had the potential to positively impact on the client's target difficulties and these techniques were appropriately applied.
i. Homework	9. Between-session activities: The rationale for between-session activities was clearly linked to recovery goals.
j. Quality of intervention	10. Facilitating appropriate emotional expression: The emotional state of the client was acknowledged and normalized where appropriate 11. Therapist's use of the guide:* The therapist referred to the guide during session planning and appropriately incorporated work on the guide into the session plan. 12. Therapist's skill at delivering the guide over the telephone.* The therapist explicitly enquired about non-observable cues/reactions (emotional, reflective or otherwise non-verbal) and verbalized feedback in response (in essence, making non-observable, verbal).
Scoring: 1 = appropriately included, 0 = inappropriately omitted and 9 = appropriately omitted (9 equivalent to 1)	

*Subscales that are specific to therapy sessions conducted over the phone and for therapy sessions that incorporate the use of a self-help guide for clients

The scoring system of the ROSTA was derived from the CTS-Psy (Haddock et al., 2001). Accordingly, scores of 1 (appropriately included), 0 (inappropriately omitted) or 9 (appropriately omitted; 9 converted to 1 for scoring purposes) are allocated to each item, allowing totals of each subscale to be calculated. An example of appropriate omission might be that session planning be omitted in circumstances that warrant abandoning the standard format (e.g. the service user is so distressed that they are unable to participate in planning how to use the session). The whole scale consists of 12 subscales; 5 of these contribute to the General section and rate general therapeutic skills necessary for the delivery of CBT and 7 contribute to the Specific section, which rates therapist skill in delivering the specific elements of the therapy. The maximum score for the whole scale is 79 (this comprises the sum of all the subscale totals), with a maximum score of 34 on the General section (this comprises the sum of all the general subscales) and 45 on the Specific section (this comprises the sum of all the specific subscales).

In addition to the subscales, an additional section on specific recovery-focused CBTp techniques was incorporated. This section lists various possible CBTp strategies (e.g. assessment, formulation, reality monitoring, behavioural experiments). This is to allow scale users to record the content of a therapy session (useful for supervision and research purposes) and facilitates the scoring of subscale 8, "choice of intervention"; enabling raters to highlight which techniques were utilized within the session and to use this to help confirm that cognitive behavioural strategies were employed appropriately and satisfactorily.

The final version of the scale was reviewed by a Service User Reference Group (SURG) comprising of people with lived experience of mental health problems (including psychosis) and who were acting as consultants to various research projects within the institutions to which the authors are affiliated. Members of this group reviewed the scale and some minor changes to the wording of items were made as a result of their recommendations. SURG provided positive feedback on the face validity of the scale, from a service user perspective.

The therapy fidelity tool is intended to be used alongside a detailed scoring manual, which provides clarification on some of the subscales and items, and details the scoring structure (this is available from the last author).

Phase 2: Psychometric evaluation

Design. Both the reliability and validity of the ROSTA scale were assessed. In terms of reliability, the internal consistency and the inter-rater reliability were investigated in relation to both the full scale and the constituent subscales. Concurrent validity assessment involved examining the degree of convergence between the ROSTA and the CTS-Psy (Haddock et al., 2001).

In all cases, investigations were undertaken that reflect the multifaceted nature of the scale, and therefore analyses were repeated excluding the two intervention-specific subscales (11 and 12) that relate to the use of the self-help guide and delivery of therapy over the telephone, to enable an understanding of the psychometric properties of the scale when used for different purposes and thus promote flexibility in its use.

Therapists and raters

Five CBT therapists, three male and two female (JK, JM, SN, CT, MW), were selected from the Division of Clinical Psychology at the University of Manchester and Greater Manchester

West Mental Health NHS Foundation Trust, to rate the therapy sessions using the ROSTA scale. All raters met the minimum training standards for the practice of cognitive behavioural therapies from the British Association of Behavioural and Cognitive Psychotherapies and had experience of working with people experiencing psychosis using CBTp delivered over the telephone and supported by a self-help guide. All raters had received training on the ROSTA scale and had engaged in several practice sessions in its use. An experienced CBTp therapist (SB) met the same standards, but was not involved in the development of the ROSTA scale and was skilled in the use of the CTS-Psy rated sessions on the CTS-Psy for concurrent-validity purposes. SB was blind to the ROSTA scale scores.

Measures

ROSTA scale. The measure assessed in this study was the ROSTA scale, the development and content of which is described above and the subscales are displayed in Table 1. Information on the items that comprise these subscales is available from the last author. Raters were provided with a full copy of the relevant scale along with a set of rating guidelines and a set of blank score sheets that they used to score sessions.

CTS-Psy (Haddock et al., 2001). Concurrent validity was assessed by comparison of scores with the CTS-Psy (Haddock et al., 2001). The CTS-Psy is split into two sections, general and specific. The general section consists of five items designed to measure general therapy skills, required to deliver CBT but not wholly unique to CBT, including understanding, feedback and collaboration. The specific section is designed to rate skills commonly agreed to be intrinsic to CBT and consists of five items, including guided discovery and homework. The scale has been shown to have almost perfect inter-rater reliability (intraclass correlation coefficient .94) with an average of .73 across the 10 items (Haddock et al., 2001). Findings indicated that raters using the CTS-Psy can distinguish between therapists who had received CBT training and those who had not, specific therapy skills as well as general competence. A summary outlining the items that comprise the CTS-Psy scale along with an indication of how both scales are scored is available in Table 1. The rater was given a full copy of the relevant scale along with a set of rating guidelines and scoring sheets, which they were directed to complete when rating audio-recorded sessions.

Generation of recorded therapy sessions

The assessment of the psychometric properties of the ROSTA scale utilized audio-recorded therapy sessions. These therapy sessions were conducted and recorded via the telephone as part of an on-going NIHR-funded trial (STAR-T; ISRCTN50487713). Participants involved in the trial met the following inclusion criteria:

1. An ICD-10 diagnosis for non-affective psychosis, including schizophrenia, schizophreniform disorder, schizoaffective disorder and delusional disorder.
2. Aged between 18–65 years.
3. At least one month of stabilization if the person has experienced symptom exacerbation in the last 6 months.
4. Able to provide written informed consent.
5. Able to read and converse in English.

All participants whose sessions were used for the current study also provided informed consent for their therapy sessions to be recorded over the telephone.

Therapy sessions

Forty-one audio files were collated from four participating therapists and systematically reviewed against a set of criteria that ensured sessions were standardized. The criteria were as follows:

1. Therapy sessions were at least 20 minutes long.
2. The session was not the first or last session of therapy, nor was it a session involving other personnel such as key workers or family members.
3. The audio recording was of sufficient quality that both the therapist and client were clearly audible.

An inventory was created listing relevant details of each session such as therapist identifier, client identifier, date of session/session number, and length of session. Those sessions that met criteria were then subject to random selection to produce the pool of eligible recordings to be used as part of the current study. Sessions to be rated were allocated to ensure that therapists did not rate their own sessions.

Procedure

Reliability. A random sample of 10 sessions was selected for investigation of internal consistency and inter-rater reliability of the ROSTA scale. The sample featured four clients and three therapists. Each session file was provided to the five assessors (JK, JM, SN, CT and MW). The use of five assessors to rate sessions for inter-rater reliability testing was deemed sufficient, based on a power analysis that demonstrated that with a significance level of 0.05, the use of five raters scoring on 10 sessions has more than 80% power to show that an intraclass correlation is greater than .6 when its true value is .8. Each assessor listened to and rated audio-recorded therapy sessions independently.

Validity

Scores from the 10 sessions used to assess reliability were also used to assess concurrent validity, along with an additional 15 session audio recordings selected at random from the same pool. A power calculation suggested that a .05 one-sided Fisher's z test of the null hypothesis that the Pearson correlation coefficient = .5, has 80% power to detect a correlation of .8 when the sample size is 24 - therefore 25 sessions was deemed acceptable. All 25 sessions were rated on both the ROSTA scale and the CTS-Psy, by different raters, as outlined above.

Statistical analyses

Scale fine-tuning. All data were analysed using SPSS (version 16). Ratings were added to the database and were initially scrutinized for low endorsements. This involved manually examining all five raters' scores for each item across the 10 sessions. Frequency of endorsement was then calculated as a percentage by dividing the number of times each item

was given a score of 1 by the maximum number of times it could have been given a score of 1 (in this case 50), and multiplying by 100. It was decided a priori that items with very low endorsements of $\leq 20\%$ across all sessions would be removed from the analysis and would potentially be removed from the scale. Additionally, items with endorsements of $< 50\%$ across all sessions were also of interest. Any items identified to have endorsement rates below this criterion were then discussed, with regards to the discriminant power and value of these items within the scale, and required a final decision to be made as to whether any items should be eliminated or modified for future use. Internal consistency of the scale, as well as that of the two sections within the ROSTA scale (general and specific) was computed using Cronbach's alpha coefficient (Cronbach, 1951) on the item scores. This was used to determine how well the items and subscales of the scale perform as a whole, how closely related the items and subscales within each section are as a group, and whether the two sections are in fact distinct from each other.

Inter-rater reliability of the final scale

The next stage of the analysis involved using the intraclass correlation coefficient with a 95% confidence interval (ICC; Shrout and Fleiss, 1979) to measure the agreement between the five raters on the scores from the whole scale across the 10 sessions. Free-marginal multirater kappa coefficient (multirater kfree) was used to assess the inter-rater reliability amongst the raters for each of the 12 individual subscales. Multirater kfree was chosen for this particular analysis due to the limitations of using ICC when variance is low, as would be expected for the ROSTA scale, and the restrictions of the distribution of ratings across categories assumed by Fleiss' kappa (Randolph, 2005).

Validity of the final scale

A correlational analysis method was used to assess the association between total scores on the ROSTA scale and CTS-Psy, as well as the subscale scores from the two scales after removal of items 11 and 12 from the ROSTA scale. Items 11 and 12 refer to the therapists' use of the guide and the therapists' skill at delivering the guide over the telephone, of which there are no comparable subscales in the CTS-Psy.

The scores of all four tests; Cronbach's alpha coefficient, ICC, multirater kfree coefficients, and Pearson's r were interpreted as: < 0 = poor agreement, $0 - .20$ = slight agreement, $.21 - .40$ = fair agreement, $.41 - .60$ = moderate agreement, $.61 - .80$ substantial agreements, and $.81 - 1$ almost perfect to perfect agreement (Landis and Koch, 1977).

Results

Frequency of endorsements

Each of the 79 items of the ROSTA Scale were rated 50 times; once per assessor across 10 sessions. Calculations revealed that none of the sub-items were found to have received very low endorsements of 20% or less over the 10 sessions. Frequency of endorsements ranged from 42–100% and two items (“The therapist asked for feedback concerning the previous session” and “The therapist asked the client if she/he anticipated any problems in carrying

Table 2. Inter-rater reliability of the items of the ROSTA scale as demonstrated by free-marginal multirater kappa values (multirater kfree)

Subscale	Item	Kappa
General:	Session plan	.58
	Providing and eliciting feedback	.18
	Understanding	.91
	Interpersonal effectiveness	.93
	Collaboration	.91
	<i>Mean</i>	.70
Specific:	Guided discovery	.95
	Focus on key cognitive behavioural elements	.95
	Choice of intervention	.93
	Between-session activities	.18
	Facilitating appropriate emotional expression	.79
	Therapists' use of the guide	.73
	Therapists' skill at delivering the guide over the telephone	.45
	<i>Mean</i>	.71

out between-session activities and worked with the client collaboratively to overcome these”) from two different subscales were found to have relatively low endorsements compared to the other items, with endorsements of 42% and 46% across the 10 sessions, respectively. However, a decision was taken to retain these items in the scale due to their reflection of essential elements of recovery-focused CBTp, although future users of the scale may wish to consider the particular relevance of these items to their own work.

Reliability

Internal consistency. Internal consistency for the total scale was very high, with a Cronbach's alpha value of .936. The five subscales and 34 items of the General section had moderate internal consistency, with a Cronbach's alpha of .604; the seven subscales and 45 items of the Specific section had considerably better internal consistency, with a Cronbach's alpha of .864. Reviewing the whole scale without either of the items reflecting use of the Recovery guide (item 11) and telephone delivery (item 12), the Cronbach's alpha is .902; with just item 11 this is .913 and with just item 12, .928.

Inter-rater reliability. Scores from the five raters across the 10 sessions showed that the inter-rater reliability of the total scale was good, with an ICC of .738 (95% CI: .676 – .795). Reviewing the whole scale without either of the items reflecting use of the Recovery Guide (item 11) and telephone delivery (item 12), the ICC was .640 (95% CI: .588 – .718); with just item 11 this was .671 (95% CI: .597 – .741) and with just item 12, .713 (95% CI: .644 – .776).

As can be seen from Table 2, kappa values for the inter-rater reliability of the five raters for each subscale ranged from .18 to .95, with similar averages for the general and specific sections; .70 and .71, respectively. Examination of the kappa values for each subscale indicated that half of them had almost perfect inter-rater agreement with values ranging from .91 to .95, two had substantial agreement, two had moderate agreement, and two subscales had poor agreement (“providing and eliciting feedback”; “between-session activities”).

Validity

The ROSTA Scale total score correlated moderately and significantly ($r = .651, p < .01$) with the CTS-Psy total. The general section of the ROSTA correlated weakly and non-significantly with the equivalent section of the CTS-Psy ($r = .274, p = .186$), whereas the two specific sections correlated moderately and significantly (after the removal of items 11 and 12 from the ROSTA, as explained above; $r = .550, p < .01$).

Discussion

The results confirm that the newly developed ROSTA scale is generally a reliable and valid tool for assessing fidelity to recovery-focused CBTp delivered over the telephone supported by a self-help recovery guide. Service user input suggests high face validity for the scale and, promisingly, the frequency of endorsement indicated that all of the 79 items were well utilized. The two items that received relatively low endorsements of 42% and 46% were “the therapist asked for feedback concerning the previous session” and “the therapist asked the client if she/he anticipated any problems in carrying out between session activities and worked with the client collaboratively to overcome these”. Given that these items were only endorsed 21 and 23 times respectively out of a possible 50, this may indicate that further examination and modification is required in order to ensure that they encompass the essential and/or demonstrable features of feedback and between-session activities. Alternatively, their low endorsement may reflect the difficulty in assessing these items without the contextual information that familiarity with the client and/or therapeutic relationship might provide, as would be available if they are utilized as part of the supervisory process.

Reliability

The Cronbach’s alpha coefficient for the total scale was very high (.936), indicating that the ROSTA scale has strong internal consistency as predicted, and the items are operating as a cohesive construct measuring the same underlying elements of the therapeutic intervention. Internal consistency of the general section was acceptable but lower than expected with only a moderate alpha value (.604). This is in contrast to the high internal consistency seen in the specific scale (.864). The lower internal consistency seen in the general section suggests that the subscales and items are operating more independently from each other than anticipated, signifying a possible area for improvement in terms of using the scale as a single instrument rather than two separate (general and specific) sections. Alphas for the whole scale with and without each of the intervention-specific components (items 11 and 12) ranged from .90 – .93, indicating that the scale’s coherence in these different compositions is also high (see Table 2).

The inter-rater reliability coefficient for the ROSTA total scale was moderately high with an ICC of .738 and supported the hypothesis outlined. Although not as high as that of the CTS-Psy (Haddock et al., 2001), this value still compares favourably with ICCs reported in psychometric investigations of other CBT fidelity scales, such as the CTS-R (.63; Blackburn et al., 2001). The use of a confidence interval of 95% also demonstrates that nearly all the scores for subscale inter-rater reliability fell between the range of .676 and .795, with the upper value of the range being just below the threshold for “near perfect agreement” (.81) and the lower value being within the “good” range (.61 – .80). ICCs for the whole scale with and

without each of the intervention-specific components (items 11 and 12) ranged from .64 – .71, indicating good-strong agreement.

Further examinations of inter-rater reliability (see Table 2) revealed that eight of the 12 individual subscales had good to almost perfect agreement (kappa values ranging from .73 – .95). However, the remaining four subscales appeared to be less reliable, therefore only partially supporting the predictions made. “Session plan” and “therapists’ skill at delivering the guide over the telephone” had moderate inter-rater reliability, although it is not clear why this might be so. It may be that the differences between the raters’ scores for these two items is due to individual factors in perception, for example, the raters may have given a mixture of 0s and 9s when scoring the same sub-items, i.e. their judgement of what was “inappropriately omitted” and “appropriately omitted” was different for those particular items. If this is the case, it may suggest that the scale requires more directive or comprehensive guidelines within the manual to support raters when scoring these items.

Interestingly, the two items that received relatively low endorsements; “providing and eliciting feedback” and “between-session activities” also had poor inter-rater reliability. Although each subscale contained six and seven items respectively, the relatively low endorsements of these items suggest that as well as receiving fewer scores of included or appropriately omitted in comparison to the other sub-items in the scale, there was also less agreement between the raters when scoring the item as a whole. This is an intriguing finding as both feedback and between-session activities (i.e. homework) are conceptually intrinsic to the nature of CBTp and should be expected to be included in a fidelity scale. Morrison and Barratt (2010) outlined the importance of the use of these components, demonstrating that at least 80% of an expert panel rated the use of feedback and homework as an essential ingredient of CBTp. There may be a rationale for individual users of the scale to weigh the importance of these aspects of the therapeutic intervention (and assessment thereof) against the reduced psychometric reliability of the items that seek to encapsulate them.

Validity

There were significant, moderate correlations between the ROSTA Scale and the CTS-Psy, in terms of total scale scores and the specific sections, although not for the general sections. This provides some support for the concurrent validity of the ROSTA Scale in demonstrating that it shares similarities with a previously validated measure of CBTp. The two scale totals and specific subscales did not correlate too highly (i.e. $\geq .9$), which supports the contention that the ROSTA is a unique scale reflecting different aspects of therapy to the CTS-Psy. The specific section in particular has been designed to encompass key elements of CBTp requiring a focus on recovery. Therefore a moderate but highly significant correlation between the two specific sections supports the uniqueness of the scale. In contrast, the correlation between the scores on the general section was not as high as expected nor was it significant. It is possible that the focus on recovery and self-help (which is integrated into the general section on the ROSTA) has resulted in greater divergence from CBTp than was anticipated.

Limitations

Despite the generally positive findings, there are several limitations of the present study and further investigation of the psychometric properties of the ROSTA Scale is recommended.

First, it could be argued that the use of five and six raters respectively in scoring 10 sessions for reliability and 25 sessions for validity is still a relatively small number, although recommendations of the a priori power calculations were adhered to. Despite this, the authors acknowledge that using a small number of raters can result in variance lower than expected, giving a smaller range of scores across the scale, and amplifying the weight of individual endorsements in the analysis, which may lead to skewed results. Further replications of the study should also involve a larger sample of raters and therapy sessions that are independent from the raters themselves in terms of familiarity to the therapist and/or client, in order to control for the potential biases that may result from this (Rollinson et al., 2008).

In addition, the therapy sessions rated in this investigation were taken from a research trial evaluating a new delivery mode of CBTp, which employed highly skilled and experienced therapists, potentially skewing the scoring of some items toward the upper end of the range. Although raters in this study were not the therapists featured in the sessions, familiarity of the raters to the therapists and, in some cases, the clients too, must be taken into account as this may have influenced the scoring. The raters featured in this study were instrumental in the development of the ROSTA scale, which involved attending the same supervision meetings, and rating and reviewing therapy audio files using the scale at various stages of its progress. Again, this may have influenced the way in which raters scored the sessions, and could have led to overly positive, unrepresentative results.

This study only examined face and concurrent validity; however, there are many other types of validity that can also be assessed. These include the investigation of content validity, established by expert judgements, and discriminant construct validity, established by assessing the degree to which the scale correlates with existing measures of other constructs. It may also be pertinent to explore the extent to which adherence relates to change denoted by subjective recovery and symptom measures, such as the QPR (Neil et al., 2009) and the SEPS (Haddock et al., 2011).

Finally, the therapeutic intervention that the ROSTA scale relates to has not yet been fully evaluated and so the utility of this new tool may only become fully apparent once the efficacy of the therapy it promotes adherence to is established.

Applications of the ROSTA scale

There are a number of possible applications of the ROSTA scale, for clinical practice, research and training. With the increased relevance of finding alternatives to face-to-face CBT (Shapiro et al., 2003), it is likely that interest in and demand for the use of a stepped-care approach to the CBTp will also increase. The scale, incorporating both general recovery-oriented CBT for people with experience of psychosis, the use of a self-help guide and telephone delivery can therefore be used in various capacities: for example, to promote therapist competence in and adherence to recovery-focused therapy delivered in a number of ways, depending on the resources of the service, expertise of the practitioners and preferences of the clients.

A meaningful use of the ROSTA scale may be as a tool involved in the development and evaluation of training programmes, and in facilitating reflective practice in supervision of new therapists training in this way of delivering CBTp. The scale may also provide a means of monitoring adherence and competence during treatment to ensure clients are receiving high-quality therapy. Subscales 11 and 12 provide a good “checklist” to ensure therapists are implementing the approach correctly over the phone and involving self-help materials in

promoting clients' recovery. The removal of items 11 and/or 12 from the scale would also provide a framework for assessing fidelity to recovery focused CBTp delivered face-to-face, further expanding the use of the ROSTA scale.

Given the current emphasis on recovery oriented practice and the current lack of research in this area, it is anticipated that future research trials may focus on this way of delivering CBTp in order to add to the evidence base in this area. Therefore, we suggest that the ROSTA scale could be useful in other studies evaluating CBTp.

Conclusion

The results of the psychometric investigations of the ROSTA scale have shown that it is a practical, easy to use scale for evaluating fidelity to recovery-oriented CBTp. The reliability of the scale was generally good (although with some notable exceptions in terms of single items) and the moderate correlation with the CTS-Psy (Haddock et al., 2001) as a whole by in large supports the concurrent validity of the scale. The scale covers distinct and important areas of therapy and has demonstrated that it is a sensitive measure of fidelity. Taken together, the findings from this study indicate that the scale could potentially make a valuable contribution towards future research involving CBTp, recovery in psychosis, telephone-based interventions and interventions that incorporate the use of a self-help guide and be utilized effectively in clinical trials.

Acknowledgments

This article presents data collected as part of the STAR Trial (ISRCTN50487713), which is part of the Recovery programme that constitutes independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research scheme (RP-PG-0606–1086). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Acknowledgements are due to individual members of the Service User Reference Group, Yvonne Awenat, Rory Byrne, Ellen Hodson, Sam Omar, Liz Pitt, Jason Price, Tim Rawcliffe and Yvonne Thomas for their work on this study. We would also like to thank a service user-researcher for assistance with understanding clients' subjective experience of CBT. Finally, we would like to thank Katherine Berry who provided some of the audio files for rating.

References

- Anthony, W. A.** (1993). Recovery from mental illness: the guiding vision of the mental health service system in the 1990s. *Psychosocial Rehabilitation Journal*, 16, 11–23.
- Bell, K., Startup, M., French, P., Morrison, A., Bucci, S. and Fowler, D.** (2008). Therapist treatment adherence in CBT for people at ultra-high risk for psychosis: development of a new rating scale. *Early Intervention in Psychiatry*, 2, A16–A16.
- Bell, A., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D., Ory, M., et al.** (2004). Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH behavior change consortium. *Health Psychology*, 23, 443–451. doi: 10.1037/0278-6133.23.5.443
- Blackburn, I.-M., James, I. A., Milne, D. L., Baker, C., Standart, S., Garland, A., et al.** (2001). The Revised Cognitive Therapy Scale (CTS-R): psychometric properties. *Behavioural and Cognitive Psychotherapy*, 29, 431–446.

- Coleman, R.** (1999). *Recovery: an alien concept*. Gloucester: Handsell Press.
- Haddock, G., Devane, S., Bradshaw, T., McGovern, J., Tarrier, N., Kinderman, P., et al.** (2001). An investigation into the psychometric properties of the cognitive therapy scale for psychosis (CTS-Psy). *Behavioural and Cognitive Psychotherapy*, 29, 221–233. doi: 10.1017/S1352465801002089
- Haddock, G., Wood, L., Watts, R., Dunn, G., Morrison, A. P. and Price, J.** (2011). The Subjective Experiences of Psychosis Scale (SEPS): Psychometric evaluation of a scale to assess outcome in psychosis. *Schizophrenia Research*, 133, 244–249. doi: 10.1016/j.schres.2011.09.023
- Keeley, H., Williams, C. and Shapiro, D.** (2002). A United Kingdom survey of accredited cognitive behaviour therapists' attitudes towards and use of structured self-help materials. *Behavioural and Cognitive Psychotherapy*, 30, 193–203. doi: 10.1017/S1352465802002060
- Landis, J. and Koch, G.** (1977). Measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Lovell, K., Cox, D., Haddock, G., Jones, C., Raines, D., Garvey, R., et al.** (2006). Telephone administered cognitive behavioural therapy for treatment of obsessive compulsive disorder: randomised controlled non-inferiority trial. *The British Journal of Psychiatry*, 333, 883–886.
- Lovell, K. and Richards, D.** (2000). Multiple access points and levels of entry (MAPLE): ensuring choice, accessibility and equity for CBT services. *Behavioural and Cognitive Psychotherapy*, 28, 379–392.
- May, R.** (2004). Making sense of psychotic experiences and working towards recovery. In J. F. Gleeson and P. D. McGorry (Eds.), *Psychological Interventions in Early Psychosis: a treatment handbook*. Chichester: Wiley.
- Mohr, D., Hart, S., Julian, L., Catledge, C., Honos-Webb, L., Vella, L., et al.** (2005). Telephone-administered psychotherapy for depression. *Archives of General Psychiatry*, 62, 1007–1014.
- Moncher, F. and Prinz, R.** (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, 11, 247–266.
- Morrison, A. P. and Barratt, S.** (2010). What are the components of CBT for psychosis? A Delphi study. *Schizophrenia Bulletin*, 36, 136–142.
- National Institute for Clinical Excellence** (2009). *Core Interventions in the Treatment and Management of Schizophrenia in Adults in Primary and Secondary Care*. CG82. London: NICE.
- Neil, S. T., Kilbride, M., Pitt, L., Nothard, S., Welford, M., Sellwood, W., et al.** (2009). The questionnaire about the process of recovery (QPR): a measurement tool developed in collaboration with service users. *Psychosis*, 1, 145–155. doi: 10.1080/17522430902913450
- Pitt, L., Kilbride, M., Nothard, S., Welford, M. and Morrison, A. P.** (2007). Researching recovery from psychosis: a user-led project. *Psychiatric Bulletin*, 31, 55–60.
- Randolph, J. J.** (2005). *Free-Marginal Multirater Kappa (Multirater Free): an alternative to Fleiss fixed-rater kappa*. Paper presented at the Joensuu University Learning and Instruction Symposium, Joensuu, Finland.
- Rollinson, R., Smith, B., Steel, C., Jolley, S., Onwumere, J., Garety, P., et al.** (2008). Measuring adherence in CBT for psychosis: a psychometric analysis of an adherence scale. *Behavioural and Cognitive Psychotherapy*, 36, 163–178. doi: 10.1017/S1352465807003980
- Sarin, F., Wallin, L. and Widerlov, B.** (2011). Cognitive behavior therapy for schizophrenia: a meta-analytical review of randomized controlled trials. *Nordic Journal of Psychiatry*, 65, 162–174. doi:10.3109/08039488.2011.577188
- Shapiro, D., Cavanagh, K. and Lomas, H.** (2003). Geographic inequity in the availability of cognitive behavioural therapy in England and Wales. *Behavioural and Cognitive Psychotherapy*, 31, 185–192. doi: 10.1017/S1352465803002066
- Shrout, P. E. and Fleiss, J. L.** (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428.

- Startup, M., Jackson, M. and Pearce, E.** (2002). Assessing therapist adherence to cognitive-behaviour therapy for psychosis. *Behavioural and Cognitive Psychotherapy*, *30*, 329–339. doi:10.1017/S1352465802003077
- Stickley, T. and Wright, N.** (2011a). The British research evidence for recovery, papers published between 2006 and 2009 (inclusive). Part One: a review of the peer-reviewed literature using a systematic approach. *Journal of Psychiatric and Mental Health Nursing*, *18*, 247–256. doi: 10.1111/j.1365-2850.2010.01662.x
- Stickley, T. and Wright, N.** (2011b). The British research evidence for recovery, papers published between 2006 and 2009 (inclusive). Part Two: a review of the grey literature including book chapters and policy documents. *Journal of Psychiatric and Mental Health Nursing*, *18*, 297–307. doi: 10.1111/j.1365-2850.2010.01663.x
- Waltz, J., Addis, M., Koerner, K. and Jacobson, N.** (1993). Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, *61*, 620–630.
- Wykes, T., Steel, C., Everitt, B. and Tarrier, N.** (2008). Cognitive behavior therapy for schizophrenia: effect sizes, clinical models, and methodological rigor. *Schizophrenia Bulletin*, *34*, 523–537. doi: 10.1093/schbul/sbm114