

# Therapy Contamination as a Measure of Therapist Treatment Adherence in a Trial of Cognitive Behaviour Therapy versus Befriending for Psychosis

Sarah Bendall, Kelly Allott and Martina Jovev

*Orygen Youth Health Research Centre, and University of Melbourne, Australia*

Marie-Josée Marois

*Centre Hospitalier Universitaire de Québec, Canada*

Eoin J. Killackey

*Orygen Youth Health Research Centre, and University of Melbourne, Australia*

John F. Gleeson

*Australian Catholic University, Melbourne, Australia*

Mario Alvarez-Jimenez and Patrick D. McGorry

*Orygen Youth Health Research Centre, and University of Melbourne, Australia*

Henry J. Jackson

*University of Melbourne, Australia*

**Background:** High quality randomized controlled trials (RCT) of psychotherapeutic interventions should ensure that the therapy being tested is what is actually delivered. However, contamination of one therapy into the other, a critical component of treatment adherence, is seldom measured in psychotherapy trials of psychosis. **Aims:** The aim of the study was to determine whether a purpose-designed measure, the ACE Treatment Integrity Measure (ATIM) could detect therapy contaminations within a controlled trial of cognitive behavioural therapy (CBT) versus Befriending for first-episode psychosis and to compare

---

Reprint requests to Sarah Bendall, Orygen Youth Health Research Centre, Locked Bag 10, Parkville, Victoria, Australia 3052. E-mail: sbendall@unimelb.edu.au

© British Association for Behavioural and Cognitive Psychotherapies 2013

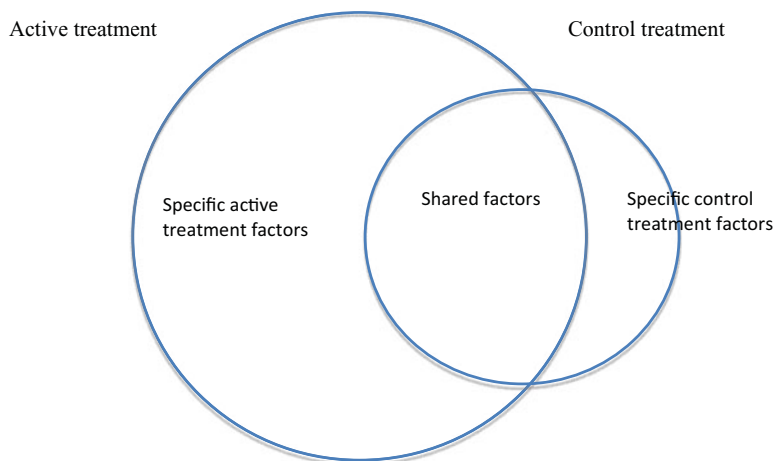
the ATIM to a more traditional adherence measure, the Cognitive Therapy Scale (CTS). **Method:** Therapy sessions were audio-recorded and at least one therapy session from 53 of the 62 participants in the RCT was rated by an independent rater using the CTS and ATIM. **Results:** Ninety-nine therapy sessions were rated. All Befriending sessions and all but three CBT sessions were correctly identified. The ATIM showed that 29 of the 99 (29%) sessions were contaminated by techniques from the other therapy. Within the CBT sessions, 19 of the 51 sessions (37%) were contaminated by one or more Befriending techniques. Of the Befriending sessions, 10 of 48 (21%) were contaminated by ACE techniques. The mean CTS score was higher in the CBT than the Befriending group. **Conclusions:** The ATIM was able to detect contaminations and revealed more meaningful, fine-grained analysis of what therapy techniques were being delivered and what contaminations occurred. The study highlights the benefit of employing purpose-designed measures that include contamination when assessing treatment adherence.

*Keywords:* Treatment adherence, treatment fidelity, cognitive behavioural therapy, first episode psychosis, contamination.

## Introduction

A high quality randomized controlled trial (RCT) of the efficacy of any psychotherapeutic intervention is expected to ensure that the therapy being tested is what is actually delivered, referred to as treatment adherence (Boutron, Moher, Altman, Schulz and Ravaud, 2008; Waltz, Addis, Koerner and Jacobson, 1993; Wykes, Steel, Everitt and Tarrrier, 2008). However, there are often important elements missing in the way adherence is measured in psychotherapy studies (Waltz et al., 1993). Many tests of treatment adherence focus on whether important or essential components of a particular therapy are present as they should be. However, it has been suggested that, in any test of adherence, it is essential to include the identification of slippage or “contamination” of one therapy into another; i.e. the presence of components that are proscribed from the therapy (Waltz et al., 1993). Following the recommendations of Waltz and colleagues (1993), a rigorous method to measure treatment adherence is to identify the major components of the therapies in question (as determined by their respective treatment manuals). The components are categorized as: 1) unique to the therapy; 2) proscribed - should not be present in the therapy; and 3) common to all therapies (e.g. Carroll et al., 2000; Hilsenroth, Blagys, Ackerman, Bonge and Blais, 2005; Segal, Teasdale, Williams and Gemar, 2002). This method lends itself to utilization in trials comparing two active psychotherapeutic treatments. However, the method may also be applied to trials employing a control psychotherapy treatment. This requires conceptualizing the control treatment as more than the non-specific elements of an “active therapy”, such as cognitive behavioural therapy (CBT). The control treatment should be manualized and have a rationale of its own in order to match client expectancies (Bendall et al., 2006). This conceptualization of the relationship between an active and control treatment is depicted in Figure 1. While there are overlapping “non-specific” factors common to both therapies, each has its own unique factors, with the active treatment having a greater number and sophistication of these. As psychotherapeutic treatments for psychosis are still relatively new, many studies utilize a psychotherapy control treatment rather than an active comparison (e.g. Jackson et al., 2008; Shawyer et al., 2012).

Numerous systematic reviews and meta-analyses of CBT for psychosis trials have concluded that CBT is efficacious, but fewer reviews have explored the measurement or



**Figure 1.** (Colour online) Model of the relationship of shared and specific treatment factors between an active and a control treatment in a psychotherapy controlled trial

quality of treatment adherence in the studies reviewed (e.g. Jones, Cormac, Silveira da Mota Neto and Campbell, 2004; Pilling et al., 2002; Zimmermann, Favrod, Trieu and Pomini, 2005). One exception was a recent review examining the methodological sophistication of studies of CBT for schizophrenia, which included the examination of the quality of the treatment adherence measures used (Wykes et al., 2008). Thirty-four RCTs were rated on the Clinical Trial Assessment Measure (CTAM; Tarrier and Wykes, 2004), which includes a subscale that assesses whether the therapy is manualized and adherence is measured (called Treatment Description). The maximum score on the subscale is 11 and of the 34 studies included in the review, scores ranged from 0–11, with a mean of 6.4 ( $SD = 3.5$ ; Wykes et al., 2008).

The Wykes and colleagues (2008) review showed that there has been an evolution in the measurement of treatment adherence in psychotherapy trials for psychosis. Recent studies have used purpose-designed or psychometrically validated measures (Alvarez-Jimenez et al., 2008; Rollinson et al., 2008; Sensky et al., 2000; Startup, Jackson and Pearce, 2002). However, these measures have focused on the CBT intervention and have not assessed adherence to the control treatment. In some cases this is because the trial being assessed did not involve a control treatment (Startup et al., 2002). In other cases the control treatment was assessed against criterion developed for the assessment of CBT. Low scores on the CBT measure were considered to reflect an absence of CBT techniques, therefore indicating adherence to the control treatment (Sensky et al., 2000). Only one psychotherapy trial in psychosis has investigated contaminations to our knowledge. This was an RCT of acceptance-based CBT versus Befriending for psychosis and, in an analysis of the audiotapes of 31 sessions, no contamination was found in either condition (Shawyer et al., 2012). While the investigation of contaminations has theoretical validity and is beginning to be used in RCTs in psychosis (e.g. Shawyer et al., 2012), the method is still relatively unknown in relation to either trials with control conditions or in people with psychosis.

We developed the ACE Treatment Integrity Measure (ATIM) for our previously reported RCT of a CBT intervention (Active Cognitive Therapy for Early Psychosis: ACE) versus a control treatment (Befriending) in first-episode psychosis (FEP; Jackson et al., 2008). The ATIM was developed in accordance with the recommendations of Waltz and colleagues (1993) and was designed to capture contaminations. As the Waltz et al. recommendations were developed specifically for comparison of two active therapies, we were interested to explore whether this method could be used to detect contamination in a trial with a control treatment. We also employed a second, psychometrically tested and valid treatment integrity measure - the Cognitive Therapy Scale (CTS; Vallis, Shaw and Dobson, 1986; Young and Beck, 1980) - in the trial. The CTS had been used to measure treatment adherence in a previous RCT of CBT for psychosis (Sensky et al., 2000). The aims of the study were, first to determine whether the ATIM could detect contaminations in our trial of CBT for FEP, and second to determine whether the detection of contaminations led to a more meaningful evaluation of treatment adherence compared to the more traditional approach of measuring the performance of each therapy on a CBT adherence scale (the CTS).

## Method

### *Participants and design*

The data presented here were derived from the treatment integrity ratings of audiotaped therapy sessions from the previously described RCT of individual CBT (ACE) versus Befriending delivered during the acute phase of first episode psychosis (Jackson et al., 2008). Both therapies were manualized (see Bendall, Killackey, Jackson and Gleeson, 2003; Bendall, Killackey, Marois and Jackson, 2005). ACE therapy comprised a flexible, formulation-based CBT approach focusing on risk, positive psychotic symptoms, co-morbidities and identity issues. Befriending consisted of talking about neutral topics such as music, sport, and pets and redirecting the conversation away from emotionally-loaded topics such as symptoms. Befriending was found to match ACE on measures of participant expectations, positive experiences of therapy and drop-out rate (Bendall et al., 2006). Participants received usual background treatment within a specialized FEP service (i.e. comprehensive case management, medication, group program) in addition to ACE or Befriending. Participants were randomly allocated to one of two experienced clinical psychologists (SB, EK) who delivered both treatments. The therapists received 3 months training in the two therapies, which included reading and synthesizing applicable therapy manuals and the completion of pre-trial pilot therapy cases. The training and trial therapy was supervised by a senior clinical psychologist (JG). Both therapists treated similar numbers of participants in both conditions. The trial protocol specified that each participant could receive up to 20 sessions of individual therapy within 14 weeks.

The participants were 62 FEP patients (31 randomized to each therapy condition) aged 16–28 years ( $M = 22.3$  years,  $SD = 3.5$ ) and are described elsewhere (Jackson et al., 2008). The study was conducted at the Early Psychosis Prevention and Intervention Centre (EPPIC), a government-funded service within Orygen Youth Health (Melbourne, Australia) that treats young people experiencing FEP (including affective and non-affective psychoses). Ethics approval was obtained from the Northwestern Mental Health Behavioural and Psychiatric

Research and Ethics Committee and written informed consent to participate in the study was obtained from all participants. This included having their therapy sessions audiotaped.

### *Measurement of therapist treatment adherence*

Therapist treatment adherence was measured using two rating scales: the ATIM and the CTS (Young and Beck, 1980). Most therapy sessions were audiotaped using a minidisk recorder and then transferred onto compact disk for storage. Of the 486 therapy sessions that were conducted in the trial, 127 sessions were not recorded or stored due to a) four participants declining consent to have sessions recorded (15 sessions) and b) operator or equipment error (112 sessions), resulting in a total of 359 recorded sessions (74%) from 53 of the 62 participants. For each participant with available data, up to four therapy sessions were selected for treatment integrity rating: one session from the first five sessions was randomly selected, and then one from each subsequent block of five. This ensured that different stages of therapy were represented in the assessment of treatment integrity (see Waltz et al., 1993). A total of one therapy session was rated for 25 participants, 2 were rated for 13 participants, 3 were rated for 12 participants and 4 were rated for 3 participants, depending on how many sessions they received in total (e.g. if a participant received 12 sessions of therapy, 3 were rated: one from sessions 1–5; one from sessions 5–10; and then one from either session 11 or 12). An independent clinical psychology doctoral student (MJ), blind to treatment allocation, listened to and rated 99 (51 ACE, 48 Befriending) sessions using both the CTS and the ATIM. A second independent clinical psychology doctoral student (MJM) re-rated 47 (24 ACE, 23 Befriending) sessions to assess the inter-rater reliability of the two measures.

### *The ATIM*

We developed the ATIM (Appendix A) to economically and reliably measure therapist treatment adherence, including therapy contamination, in line with the guidelines described by Waltz et al. (1993) (Figure 1). In conjunction with development of the respective therapy manuals (Bendall et al., 2003, 2005), components of ACE and Befriending that were central and unique to each therapy were generated by consensus between the treating psychologists and randomly listed in the ATIM. The 26 items comprised two subscales reflecting ACE (21 items) and Befriending (5 items), respectively. For example, a sample item pertaining to ACE states: “Therapist guides client in problem solving” and a sample item relating to Befriending states: “Therapist chooses most neutral line of questioning”. Raters record whether each of the 26 items (21 ACE items, 5 Befriending items) was present or absent (rated as 1 or 0) in the therapy session. The ATIM also has a third subscale, called General Therapy Techniques, comprising four items, rated as present or absent, which are techniques mutual to the two therapies and not unique to one specific therapy (e.g. Empathy). Finally, a single question (31) on the ATIM scale asks “Which form of therapy is being done?” in order to determine whether raters could broadly classify the session as either ACE or Befriending.

We divided the analysis of the ATIM into four areas: broad therapy classification; adherence; contamination; and general therapy techniques. Broad therapy classification was measured by the final question of the ATIM, which asked raters to classify each therapy session as either ACE or Befriending. To check for adherence (i.e. whether the therapy that was meant to be conducted was actually present), we set a minimum criterion that at least one

item from the subscale of the therapy in question was checked as present (i.e. a minimum score of 1 on the subscale in question). If no item was present for the intended therapy (i.e. a score of 0 on the subscale in question), then this would indicate poor treatment adherence. To check for contamination, the score on the opposing subscale was calculated (i.e. if ACE was delivered, then the score on the Befriending subscale would determine if contamination had occurred). A score of zero would indicate no contamination was detected and a score of 1 or more would indicate the therapy in question was contaminated by techniques from the opposing therapy (i.e. proscribed techniques). We expected ratings of general therapy techniques to be similar across groups as we considered that these were aspects of the common factors of all therapies and would further ensure that Befriending was a therapy able to control for these common factors (Bendall et al., 2006).

### *The CTS*

The CTS is a scale designed for the measurement of cognitive therapy competence in cognitive therapy training (Vallis et al., 1986; Young and Beck, 1980). Raters score the therapy session using 13 questions comprising three key areas of cognitive therapy: general interview procedure (e.g. Agenda setting), interpersonal effectiveness (e.g. Professionalism) and specific cognitive-behavioural techniques (e.g. Use of guided discovery). Each item is rated from 0 to 6, with higher scores indicating a more competent delivery of the particular therapy technique in question. In the current study, the total CTS score was used to measure the degree and skilfulness in which CBT was delivered in both treatment groups as per Sensky et al. (2000). Sensky et al. (2000) set a cut-off point below which Befriending therapy was expected to score (24), and above which CBT was expected to score (39) to indicate the likelihood that each therapy was delivered.

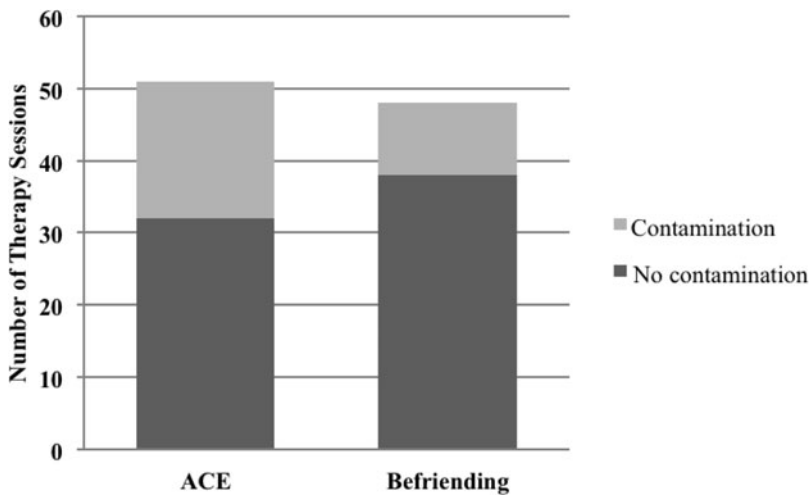
### *Statistical analyses*

We calculated descriptive statistics for the unique, proscribed, and common elements of both ACE and Befriending within the ATIM. Independent *t*-tests were used to examine differences between treatment integrity ratings of ACE and Befriending on the CTS. Cohen's kappa and intraclass correlations were used to determine inter-rater reliability on the ATIM and CTS. Analyses were conducted using SPSS for Windows (version 18.0).

## **Results**

### *Broad therapy classification*

The independent rater correctly classified 96 of 99 (97%) therapy sessions as the appropriate therapy (question 31 on the ATIM). The three incorrect classifications were rated as Befriending when they were in fact ACE. Inter-rater agreement of the broad therapy classification was high (Kappa = 0.92).



**Figure 2.** Frequency of therapy contamination of 51 ACE and 49 Befriending sessions rated on the ATIM

#### ATIM

*Adherence (unique items).* All rated ACE sessions (100%) scored a minimum of one on the ACE subscale. The mean number of CBT items endorsed in the ACE sessions was 7.71 ( $SD = 3.79$ ) out of a possible 21 items. All rated Befriending sessions (100%) scored one or more on the Befriending subscale. The mean number of Befriending items endorsed in the Befriending sessions was 2.52 ( $SD = 1.22$ ) out of a possible 5 items. This indicated that at least one unique element of the appropriate therapy was delivered as required and in both therapies therapists generally used more than one unique element of the appropriate therapy.

*Contamination (proscribed items).* Ratings on the ATIM measure showed that 19 of the 51 sessions (37%) of ACE therapy were contaminated by one or more Befriending techniques (see Figure 2). Of the Befriending sessions, 10 of 48 (21%) were contaminated by ACE techniques (see Figure 2). Of these 29 contaminated sessions, 26 (90%) were correctly identified by the external rater. Table 1 shows details of which and when specific proscribed therapy techniques occurred (i.e. therapy contamination).

At the individual person level, within ACE, for seven clients contamination occurred only once across the therapy sessions analysed; for three clients there were two contaminations, and in two cases there were three contaminations. Within Befriending, for seven clients contamination occurred only once across the sessions analysed, and in two cases contamination occurred twice.

*General therapy techniques (common items).* “Collaboration” was assessed as present in 43 of 50 CBT sessions and 33 of 47 Befriending sessions. “Empathy” was assessed as present in 48 of 50 CBT sessions and 45 of 47 Befriending sessions. “Professionalism” was assessed as present in 49 of 50 CBT sessions and 47 of 47 Befriending sessions. “Therapist direction

**Table 1.** ATIM ratings of frequency and type of therapy contamination, and the case in which contamination occurred

Therapy	Proscribed item	Frequency	Session number where contamination occurred
ACE	Therapist chooses most neutral line of questioning	13	1 <sup>d</sup> ,1,3 <sup>c</sup> ,3 <sup>e</sup> ,7 <sup>c</sup> ,9 <sup>a</sup> ,9 <sup>d</sup> ,9,9,10 <sup>e</sup> ,11 <sup>c</sup> ,11 <sup>e</sup> ,15 <sup>b</sup>
	Therapist reacts minimally to loaded speech (e.g. symptoms, conflicts)	9	1 <sup>d</sup> ,1,1,1,3 <sup>b</sup> ,3,9,14 <sup>a</sup> ,16
	Therapist redirects from discussion about symptoms to a neutral topic	1	14 <sup>a</sup>
	Therapist redirects from unresolved conflict to a neutral topic	1	9
Befriending	Therapist gives client information about the therapy process	9	1,1,1,2 <sup>f</sup> ,2,4 <sup>g</sup> ,5,6 <sup>f</sup> ,8 <sup>g</sup>
	Therapist engages the client in relaxation training	1	13
	Therapist and client work on behavioural coping strategies for specific symptoms	1	13

*Notes:* Some sessions had more than one occurrence of contamination. The same superscripts denote contaminations within an individual client.

of the therapeutic process” was assessed as present in 45 of 50 CBT therapy sessions and four of 47 Befriending sessions.

*Inter-rater reliability.* The intraclass correlation coefficients for the ATIM were as follows: ACE 0.74; Befriending 0.60.

*CTS.* The mean total CTS score was higher for the ACE group ( $M = 38.5$ ,  $SD = 13.5$ ;  $\min = 4$ ;  $\max = 65$ ) than for the Befriending group ( $M = 15.9$ ,  $SD = 3.5$ ;  $\min = 14$ ;  $\max = 27$ ) ( $p < .05$ ). As many of the 99 rated sessions were not independent of each other (as they were therapy sessions involving the same participant), the  $t$ -test analysis was repeated for a subset of 52 independent sessions (if more than one therapy session had been rated for an individual, one was randomly chosen for the analysis). The mean difference remained significant, and represented a large effect ( $r = 0.77$ ). The intraclass correlation coefficient for the total CTS score was 0.69.

## Discussion

The external rater was able to identify all Befriending sessions correctly and all but three ACE sessions correctly. This is a very gross measure of treatment adherence as it is possible that a therapy session could be contaminated while raters could still identify the particular therapy that was proscribed.



*ATIM*

In developing the ATIM we aimed to determine: 1) whether the therapists delivered the appropriate therapy (prescribed and common techniques) and 2) whether they delivered unique techniques from the proscribed therapy (i.e. contamination). We found that in all sessions rated, at least one technique central and unique to the appropriate therapy was delivered, indicating that the therapists did in fact deliver at least one element of the correct therapy. In any individual session therapists used an average of seven elements of CBT when delivering ACE, and two elements of Befriending when delivering Befriending. It could be argued that setting the minimum criterion of only one element of the appropriate therapy is a weak test of adherence, especially for CBT where there were 21 potential elements for the therapist to utilize. However, the ATIM items were not designed such that a higher number of items endorsed would equate to more adherent therapy. For example, the item "Therapist sets or discusses behavioural or cognitive experiments" reflects an intervention that would commonly take an entire therapy session to deliver. Further research is required to identify whether therapy sessions where one or few CBT techniques are utilized are less adherent or competent (e.g. less time is spent engaging in CBT activities and this is deemed inappropriate by raters). Waltz and colleagues (1993) note that use of few therapy techniques in a therapy session may in fact comprise high quality and adherent psychotherapy.

The most likely failure of treatment adherence in this trial would be treatment contamination; that is, a (highly trained CBT) therapist delivering proscribed ACE interventions in a Befriending therapy session due to factors such as habit or the perceived needs of the client. The ATIM revealed that this occurred on 11 occasions (10 sessions). Nine of these cases were on the item "Therapist gives client information about the therapy process". In development and piloting of Befriending, we aimed to orient the client to Befriending in a meeting before the first therapy session to ensure that there was no need to engage in any discussion of the therapy process during therapy. Piloting suggested that the client would take the therapist's lead in engaging in the Befriending intervention. The current data suggest that there are times where these process discussions must be had, particularly early in therapy (see Table 1). The two other contaminations of ACE into Befriending occurred in the same session. The trial protocol determined that ACE techniques could be applied in Befriending when the therapist believed there was clear clinical need (e.g. to manage risk, severe distress, or symptomatology). Whether this was the reason for the contamination or an error was made, the results show that: a) the ATIM is sensitive enough to detect contaminations; and b) that very few transgressions were detected, indicating good adherence to Befriending.

Unexpectedly, the ATIM revealed rather more transgressions of Befriending therapy into the ACE treatment, most of which represented two items: "Therapist chooses most neutral line of questioning" or "Therapist reacts minimally to loaded speech (e.g. symptoms, conflicts)". We were surprised at how often these were evident in therapy and conclude that there may be situations where these are acceptable techniques in CBT therapy, particularly in young people with acute psychosis where therapists may wish to lower levels of arousal within a therapy session due to an unstable mental state or, for example, to divert the focus from a strongly-held delusion. Befriending items that appear multiple times in ACE sessions may need to be removed from the ATIM in future studies. The fact that three ACE sessions were mistakenly identified as Befriending, but no Befriending sessions were identified as ACE, may also suggest that in some situations it may be therapeutically appropriate to conduct a

Befriending-type session within CBT treatment, especially with young people in the acute phase of early psychosis. This may promote engagement or provide non-threatening support if the client has acute psychotic symptoms. Indeed, recent research has indicated that Befriending may be therapeutic in some situations (Milne, Wharton, James and Turkington, 2006; Samarasekera et al., 2007). The previously published study (Shawyer et al., 2012) that used the ATIM did not report any contaminations of these or other Befriending items in their ACT therapy for people with command hallucinations. This difference may be due to the difference in severity of symptom in the groups: many of the participants in the Shawyer et al. study were in a stable phase whereas study participants were in the acute phase of illness in the current study.

Therapist behaviour in the ACE and Befriending sessions was similarly collaborative, empathetic and professional, supporting our previous work that suggested that Befriending therapy matches many of the common factors of therapy (Bendall et al., 2006). Therapists directed the therapeutic process a great deal less in Befriending than ACE. We had expected that in Befriending, therapists would have to direct the therapeutic process (in particular, to be skilful in redirecting young people to neutral topics). Our experience and the results of this item show that in fact young people were very willing to “talk about the good things” and relatively little process management was required.

### CTS

On the face of it the CTS appeared to provide good evidence of adherence to the correct therapy as there was a significant difference of large effect size between the mean total scores of the ACE and Befriending sessions. The results were broadly similar to those of Sensky et al. (2000). However, the mean score for the ACE group was 10 points lower than in the Sensky et al. trial and fell below their predetermined cut-off (39) for CBT in that trial (although it is unclear how the CTS cut-off scores were derived). The mean CTS score for the ACE group (38.5) was also between the mean CTS scores for CBT sessions for depression that had been independently rated as acceptable (47.31) and unacceptable (27.28) by expert CBT clinicians (Vallis et al., 1986). The most likely reason for our lower mean CTS score for the ACE group is that the measure is scored assuming that all CBT techniques on the scale must be delivered in all therapy sessions. In the current study of young people with acute psychosis, some therapy techniques were not appropriate for some individuals in certain sessions, whereas more therapy techniques would have been able to be delivered to people without severe acute psychotic symptoms. Another possible reason is that therapists in the ACE group did not deliver CBT as competently as in the studies described above. The design of the CTS makes it impossible to determine which of these explanations is correct, highlighting a significant flaw of the CTS as a measure of adherence for treatment of acute psychosis. Waltz and colleagues (1993) criticized treatment integrity measures in general for assuming that good quality therapy involves the same techniques, regardless of contextual factors such as phase of therapy, the age or symptom severity of clients.

This issue also leads to a problem when using the CTS as an adherence measure for any therapy other than CBT. Each item's anchors reflect how competently a technique is delivered rather than its presence or absence. For example, the anchor for the lowest score (0) on the “Application of cognitive techniques” item of the CTS states: “Therapist's application of cognitive techniques was poorly executed”. As the rater must score every item, this

anchor must be endorsed even if no cognitive techniques were present in the session. This conflates 1) a CBT session with no cognitive techniques present (but possibly others present such as engagement or behavioural interventions); 2) a CBT session with poor quality CBT techniques executed; 3) a CBT session with competently delivered Befriending techniques (i.e. contamination); and 4) a Befriending session with no CBT techniques executed. This undermines the sensitivity of the measure and its ability to capture treatment contamination. As noted previously, the most likely failure of treatment adherence in this trial is a therapist delivering ACE techniques in a Befriending therapy session. The use of a proscribed ACE technique in a Befriending session, particularly if executed briefly, partially or poorly, could have elicited the same score on the CTS as if it had not been present at all.

### *Limitations*

The inter-rater reliability of both the ATIM and CTS was less than desirable. It is possible that the reasons for this were different for each scale. The CTS items are well described and have anchors for each scale point. However, as described above, these are not always applicable, especially in the Befriending sessions, and may have resulted in the raters rating idiosyncratically. Items on the ATIM were less comprehensively described, which may have led to different ratings. We randomly sampled one session from the first five, and then one from each subsequent block of five sessions for each participant in order to control for potential differences in early, middle and late therapy sessions (Waltz et al., 1993). This however may have biased the results as more sessions of treatment completers were rated. In future a sampling approach that balances both phase of treatment (e.g. early, late) and completion of therapy may be more appropriate. All ATIM Befriending items were concerned with redirecting away from loaded topics rather than focusing on what was the main content of Befriending therapy: focusing on positive and neutral topics. In contrast, CBT items were concerned with techniques central to CBT treatment. The ATIM could be expanded to include more positive items (e.g. "discussion of topic of interest to the client"). Future research could also involve assessing clients' engagement with discussion of neutral topics as this could impact therapy retention, therapeutic engagement, and even outcome.

In conclusion, this study demonstrated that the CTS showed significant differences between the two therapy groups, possibly indicating good treatment adherence. However, it was not able to detect contamination of proscribed therapy techniques or differentiate between poor quality CBT and Befriending. The ATIM was designed to detect both proscribed techniques and contaminations, and did so. Results from the ATIM showed that, generally, there were very few contaminations of specific CBT techniques into Befriending, which was arguably the biggest threat to treatment adherence in this trial. This suggests that Befriending is a treatment that can be appropriately delivered to young people with acute psychosis (Bendall et al., 2006) and that therapists were able to deliver it without resorting to CBT techniques. The ATIM also showed that CBT techniques were able to be delivered in the CBT arm with this client group despite their having very acute symptoms. Therefore the ATIM was able to show that the therapies being tested were what was actually delivered in this trial - the ultimate aim of a treatment adherence measure. Both measures, however, showed relatively poor inter-rater reliability. Future studies using the ATIM require refinement of the measure, with a view to improving psychometric properties, including inter-rater reliability. Another future direction is to include a therapist competence component in the ATIM.

Studies of treatment fidelity in trials of CBT for psychosis have shown that, as each treatment differs in its emphasis on different CBT techniques, treatment fidelity measures must be tailored to the treatment manual (Rollinson et al., 2008). As such, each new treatment must have a new treatment fidelity measure, which means that reliability and validity can only be established retrospectively. However, as research into treatment fidelity continues, we can create tailored, valid and reliable fidelity measures based on the principles explored in previous research.

### Acknowledgements

This study was supported by funding from the Colonial Foundation, NHMRC Project Grant #145760 (H. Jackson, P. McGorry) and NHMRC Early Career Research Fellowship #1036425 (S. Bendall). All authors declare no conflict of interest. We thank the participants for their participation in the study.

### References

- Alvarez-Jimenez, M., Wade, D., Cotton, S., Gee, D., Pearce, T., Crisp, K., et al. (2008). Enhancing treatment fidelity in psychotherapy research: novel approach to measure the components of cognitive behavioural therapy for relapse prevention in first-episode psychosis. *Australian and New Zealand Journal of Psychiatry*, 42, 1013–1020.
- Bendall, S., Jackson, H. J., Killackey, E., Allott, K., Johnson, T., Harrigan, S., et al. (2006). The credibility and acceptability of befriending as a control therapy in a randomized controlled trial of cognitive behaviour therapy for acute first episode psychosis. *Behavioural and Cognitive Psychotherapy*, 34, 277–291.
- Bendall, S., Killackey, E., Jackson, H. and Gleeson, J. (2003). *Befriending Manual*. Melbourne: Orygen Youth Health Research Centre and Department of Psychology, University of Melbourne.
- Bendall, S., Killackey, E., Marois, M.-J. and Jackson, H. (2005). *ACE Manual (Active Cognitive Therapy for Early Psychosis)*. Melbourne: Orygen Youth Health Research Centre and Department of Psychology, University of Melbourne.
- Boutron, I., Moher, D., Altman, D. G., Schulz, K. F. and Ravaud, P. (2008). Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Annals of Internal Medicine*, 148, 295–309.
- Carroll, K. M., Nich, C., Sifry, R. L., Nuro, K. F., Frankforter, T. L., Ball, S. A., et al. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*, 57, 225–238.
- Hilsenroth, M. J., Blagys, M. D., Ackerman, S. J., Bonge, D. R. and Blais, M. A. (2005). Measuring psychodynamic-interpersonal and cognitive-behavioral techniques: development of the Comparative Psychotherapy Process Scale. *Psychotherapy*, 42, 340–356.
- Jackson, H. J., McGorry, P. D., Killackey, E., Bendall, S., Allott, K., Dudgeon, P., et al. (2008). Acute-phase and 1-year follow-up results of a randomized controlled trial of CBT versus Befriending for first-episode psychosis: the ACE project. *Psychological Medicine*, 38, 725–735.
- Jones, C., Cormac, I., Silveira da Mota Neto, J. I. and Campbell, C. (2004). Cognitive behaviour therapy for schizophrenia. *The Cochrane Database of Systematic Reviews*, Issue 4.
- Milne, D., Wharton, S., James, I. and Turkington, D. (2006). Befriending versus CBT for schizophrenia: a convergent and divergent fidelity check. *Behavioural and Cognitive Psychotherapy*, 34, 25–30.

- Pilling, S., Bebbington, P., Kuipers, E., Garety, P., Geddes, J., Orbach, G., et al.** (2002). Psychological treatments in schizophrenia: I. meta-analysis of family intervention and cognitive behaviour therapy. *Psychological Medicine*, *32*, 763–782.
- Rollinson, R., Smith, B., Steel, C., Jolley, S., Onwumere, J., Garety, P. A., et al.** (2008). Measuring adherence in CBT for psychosis: a psychometric analysis of an adherence scale. *Behavioural and Cognitive Psychotherapy*, *36*, 163–178.
- Samarasekera, N., Kingdon, D., Siddle, R., O’Carroll, M., Scott, J. L., Sensky, T., et al.** (2007). Befriending patients with medication-resistant schizophrenia: can psychotic symptoms predict treatment response? *Psychology and Psychotherapy: Theory Research and Practice*, *80*, 97–106.
- Segal, Z. V., Teasdale, J. D., Williams, J. M. and Gemar, M. C.** (2002). The mindfulness-based cognitive therapy adherence scale: inter-rater reliability, adherence to protocol and treatment distinctiveness. *Clinical Psychology and Psychotherapy*, *9*, 131–138.
- Sensky, T., Turkington, D., Kingdon, D., Scott, J. L., Scott, J., Siddle, R., et al.** (2000). A randomized controlled trial of cognitive-behavioral therapy for persistent symptoms in schizophrenia resistant to medication. *Archives of General Psychiatry*, *57*, 165–172.
- Shawyer, F., Farhall, J., Mackinnon, A., Trauer, T., Sims, E., Ratcliff, K., et al.** (2012). A randomised controlled trial of acceptance-based cognitive behavioural therapy for command hallucinations in psychotic disorders. *Behaviour Research and Therapy*, *50*, 110–121.
- Startup, M., Jackson, M. and Pearce, E.** (2002). Assessing therapist adherence to cognitive-behaviour therapy for psychosis. *Behavioural and Cognitive Psychotherapy*, *30*, 329–339.
- Tarrier, N. and Wykes, T.** (2004). Is there evidence that cognitive behaviour therapy is an effective treatment for schizophrenia? A cautious or cautionary tale? *Behaviour Research and Therapy*, *42*, 1377–1401.
- Vallis, T. M., Shaw, B. F. and Dobson, K. S.** (1986). The Cognitive Therapy Scale: psychometric properties. *Journal of Consulting and Clinical Psychology*, *54*, 381–385.
- Waltz, J., Addis, M. E., Koerner, K. and Jacobson, N. S.** (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.
- Young, J. E. and Beck, A. T.** (1980). *Cognitive Therapy Scale: rating manual*. Philadelphia, PA: Unpublished Manuscript, University of Pennsylvania.
- Zimmermann, G., Favrod, J., Trieu, V. H. and Pomini, V.** (2005). The effect of cognitive behavioral treatment on the positive symptoms of schizophrenia spectrum disorders: a meta-analysis. *Schizophrenia Research*, *77*, 1–9.

**Appendix A: ACE Treatment Integrity Measure (ATIM)**


---

ID number:                      Session number:                      Rater initials:                      Date:

For each item below, circle 0 if absent or 1 if present for this therapy session:

*Specific therapy techniques:*

Absent Present

0	1	1. Therapist develops a formulation with the client (e.g. makes links between past events, current situation and symptoms or relates past events, current situation and symptoms to a cognitive model)
0	1	2. Assessment/information gathering concerning symptoms, past life events
0	1	3. Therapist engages with client in setting goals for therapy
0	1	4. Therapist gives client psychoeducation
0	1	5. Therapist redirects from unresolved conflicts to a neutral topic
0	1	6. Therapist uses motivational interviewing techniques
0	1	7. Therapist sets or follows-up homework
0	1	8. Therapist redirects from discussion about symptoms to a neutral topic
0	1	9. Therapist focuses on key cognitions
0	1	10. Therapist and client work on thought records
0	1	11. Therapist and client engage in role reversal exercises
0	1	12. Therapist chooses most neutral line of questioning
0	1	13. Therapist guides the client in perspective taking
0	1	14. Therapist engages in cognitive challenging
0	1	15. Therapist reacts minimally to loaded speech (e.g. symptoms, conflicts)
0	1	16. Therapist and client work on cognitive coping strategies for specific symptoms
0	1	17. Therapist redirects from identity issues to a neutral topic
0	1	18. Therapist uses guided discovery/Socratic questioning
0	1	19. Therapist sets or discusses behavioural or cognitive experiments
0	1	20. Therapist engages in cognitive restructuring
0	1	21. Therapist guides client in identifying triggers
0	1	22. Therapist gives client education about the therapy process
0	1	23. Therapist guides client in problem solving
0	1	24. Therapist engages the client in exposure training
0	1	25. Therapist engages the client in relaxation training
0	1	26. Therapist and client work on behavioural coping strategies for specific symptoms

*General therapy techniques:*

Reflecting on the whole session, were the following in evidence?

Absent Present

0	1	27. Collaboration
0	1	28. Empathy
0	1	29. Professionalism
0	1	30. Therapist direction of the therapeutic process

31. Which form of therapy is being done?

0	Cognitive behavioural therapy
1	Befriending

---