Outcomes of patients completing and not completing cognitive therapy for depression

Jane Cahill\textsuperscript{1}, Michael Barkham\textsuperscript{1}, Gillian Hardy\textsuperscript{1}, Anne Rees\textsuperscript{1}, David A. Shapiro\textsuperscript{1}, William B. Stiles\textsuperscript{2} and Norman Macaskill\textsuperscript{1}

\textsuperscript{1}University of Leeds, UK
\textsuperscript{2}Miami University, USA

\textbf{Objectives.} The aims of this study were to use symptom intensity measures collected at each session (1) to describe the outcomes of clients who received cognitive therapy (CT) for depression in a clinically representative sample, and (2) to compare the outcomes of clients who completed the agreed number of sessions with those who did not.

\textbf{Design and method.} Clients (N = 58) contracted to attend between 12 and 20 sessions of CT completed the Beck Depression Inventory (BDI) immediately prior to each therapy session. The BDI and other measures were collected at intake and, for those who completed therapy, at a post-therapy assessment.

\textbf{Results.} Completers’ BDI scores improved significantly from intake to post-treatment and significantly more from intake to their final session than did those of non-completers. However, when non-completers’ final session scores were matched with scores of randomly selected completers at the corresponding session, the difference in improvement was not significant. A significantly higher proportion of clients who completed the agreed number of sessions achieved reliable and clinically significant change (71.4\%, 25/35), compared with just 13\% (3/23) of clients who did not.

\textbf{Conclusions.} (1) CT for depression can be effective in a clinically representative population. (2) Attrition from clinical trials may bias estimates of treatment effectiveness.

In randomized controlled trials (RCTs) of psychotherapy employing pre–post designs, clients who do not complete therapy are often replaced and excluded from analyses, whether or not they have made clinically important change (Kazdin, 1999). The

\textsuperscript{*}Requests for reports should be addressed to Jane Cahill, Psychological Therapies Research Centre, 17 Blenheim Terrace, Leeds LS2 9AR, UK (e-mail: janec@psychology.leeds.ac.uk).
findings of such studies may thus reflect treatment effects on a subgroup of persistent and well-motivated clients who complete therapy and may not be generalizable to clinical practice, in which client motivation and persistence is highly variable. For some purposes, this distortion may be addressed in intention-to-treat analyses, which include all clients initially accepted into the study. However, this strategy does not address the empirical question of the outcomes of clients who fail to complete. The present study assessed the effectiveness of time-limited cognitive therapy (CT) for depression in a routine practice setting and compared the outcomes of clients who completed their contracted number of sessions with those who did not.

It has been argued elsewhere that there is a fundamental difference between evidence derived from efficacy studies and practice-based evidence derived from routine practice (see Barkham & Mellor-Clark, 2000). The efficacy of various forms of CT has been well supported by numerous RCTs (e.g. Elkin, 1994; Hollon et al., 1992; Shapiro et al., 1994). However, there is less understanding of CT’s effectiveness and processes in clinically representative settings. Among other things, the drop-out rate is typically higher in clinical practice than in research trials. For example, rates were 13% in the Second Sheffield Psychotherapy Project (Shapiro et al., 1994) and 32% in the NIMH Treatment of Depression Collaborative Research Project (Elkin et al., 1989). In contrast, client drop-out rates at public clinics after five or fewer sessions were up to 50% (Pekakirk, 1991). Investigating the outcomes of specific brands of therapy under varying degrees of rigour is essential for assessing the generalizability of efficacy findings to practice (Shadish et al., 1997; Shadish, Navarro, Matt, & Phillips, 2000) and making policy recommendations regarding clients who unilaterally terminate therapy.

Our study was partially consistent with Stage 1 of Shadish et al.’s (1997, 2000) criteria for clinical representativeness. Although the study was carried out in a university setting, the research clinic was managed jointly with the health service; the study used usual referral routes for clients; and the study involved six clinical psychologists with regular caseloads comparable with those of other clinical psychologists working in Community Mental Health Trusts (CMHTs). However, the therapists did receive weekly supervision from highly qualified CT specialists, something not found in routine clinical practice. In contrast to RCTs, our study did not exclude and replace clients who dropped out of treatment. We gathered self-report data on symptom intensity immediately before every session using a standard measure, the Beck Depression Inventory (BDI; A. T. Beck & Steer, 1987; A. T. Beck, Steer, & Brown, 1996). Thus, we were able to estimate the progress of clients who either dropped out or failed to appear for scheduled post-treatment assessments, unlike traditional pre–post designs in which post-treatment data for non-completers are often unavailable. In addition to statistically comparing the mean scores of completers and non-completers, we applied the criteria of reliable and clinically significant change (Jacobson & Truax, 1991) to assess the extent to which CT was associated with psychometrically reliable change that moved people from the client population to the normal population.

**Method**

**Clients and selection procedures**

Referrals to the jointly managed health service and university research clinic were accepted from two sources:
the waiting lists of secondary care services, and
(2) primary care services (i.e. GP referrals).

Clients were excluded from the study if they had psychotic or manic symptoms, were misusing alcohol or drugs, or had received more than three sessions of one-to-one talking treatment within the previous 12 months. Clients had to score 18 or above on the BDI to be invited for an interview assessment and at assessment had to return a diagnosis of Major Depressive Episode (MDE) in accordance with DSM-IV (American Psychiatric Association, 1994).

Over the 2.7 year recruitment period, of the 84 clients invited for assessment, 77 (92%) attended. Of the 77 assessed, two were excluded because they did not receive a diagnosis of MDE, six because they did not receive CT, and three because they were included in a different research design, leaving an ‘intent to treat’ sample of 66. The relatively high proportion of referrals who were eligible reflected filtering that took place before clients were invited for assessment, based on detailed information on inclusion and exclusion criteria made available to GPs and other referral sources. Of the 66 accepted for treatment, eight withdrew before therapy started, leaving 58 clients, who attended 2–20 sessions each.

The mean age of the 58 clients who began treatment was 34.2 years (SD = 9.8); 45 (78%) were women. No restrictions were placed on clients’ racial or ethnic status, which was not recorded; most were white Anglo-Saxon. With regard to working status, 39 were in employment, nine were students, nine were unemployed and employment information was missing for one client. Sixteen were married or cohabiting, 42 were single, separated or divorced. A total of 45 of the clients were referred from primary care and 13 from secondary care.

**Therapists**
Six female UK-trained clinical psychologist therapists treated clients. They had between one and six years’ post-qualification clinical experience at the start of the project. All took part in a CT training programme and received weekly supervision from two experienced therapists who had specialist post-qualification training from CT training centres of excellence in the United Kingdom (MB) and United States (NM), respectively.

**Treatment and treatment contracts**
The CT approach emphasizes the provision by therapists of cognitive strategies for application by the client, including self-management procedures and cognitive restructuring. Therapy followed the protocol of A. T. Beck, Rush, Shaw, and Emery (1979), and the treatment manual was based on *Cognitive therapy: Basics and beyond* (J. S. Beck, 1995). Clients were seen weekly, as far as was possible. Each session lasted 50 min.

Clients were offered a minimum of 12 and a maximum of 20 sessions of CT for depression in a jointly managed health service and university research clinic. Of the clients, 28 (referred to hereon as the 12–20 session cohort) were contracted to receive a minimum of 12 sessions. After their twelfth session, clients negotiated how many more sessions they were to receive with their therapist, the maximum number of sessions being 20. Because of administrative changes in the clinic, the remaining 30
clients (referred to hereon as the 12-session cohort) had a different contract: They each received a clinical assessment and were contracted for 12 therapy sessions.

Of the 58 clients who began treatment, 23 failed to complete therapy. A ‘non-completer’ was defined as a client who dropped out of therapy before the twelfth session or, for clients in the 12–20 session cohort, before completing the number of sessions agreed with the therapist (as described above). A ‘completer’ was defined as a client who completed the contracted treatment—12 sessions for the 12-session cohort or the number agreed with their therapist (minimum of 12, as described above) for the 12–20 session cohort. Therapist notes from non-completers’ therapy files were examined and reasons for discontinuing therapy logged and categorized. Difficult or ambivalent feelings about therapy and practical difficulties in attending sessions were the two most common reasons for discontinuation of therapy.

**Assessment measures**

**Beck Depression Inventory (BDI)**
The BDI was used to measure the intensity of depressive symptoms. The BDI comprises 21 items, each of which presents a forced choice among four different responses and is scored from 0 to 3 to reflect the intensity of that symptom. Most of the clients (N = 44) completed the original version of the BDI (BDI-IA; A. T. Beck & Steer, 1987), while the remainder (N = 14) completed the revised version (BDI-II; A. T. Beck et al., 1996). As instructed in the manual (A. T. Beck et al., 1996) we converted the BDI-II scores to BDI-IA scores for our analyses.

**Inventory of Interpersonal Problems, Cluster C Scale (IIP-C)**
We used a 19-item inventory, the IIP-C, derived from the 127-item Inventory of Interpersonal Problems (Horowitz, Rosenberg, Baer, Ureno, & Villasenor, 1988), comprising two subscales, need for social approval (nine items, \( \alpha = .81 \)) and lack of sociability (10 items, \( \alpha = .87 \)). These subscales were based on items that distinguished between patients with a Cluster C personality disorder (PD) and all others (Pilkonis, Kim, Proietti, & Barkham, 1996). The IIP-C score was the total score on these two subscales. The subscales performed well on the original sample and on cross-validation (Pilkonis et al., 1996). It has been shown that Cluster C personality disorder often co-exists with depression (Shea, Widiger, & Klein, 1992).

**The Dysfunctional Attitudes Scale (DAS; Weissman and Beck, 1978)**
This is a measure of attitudes and beliefs hypothesized by A. T. Beck (1976) to underlie a propensity for negative thinking. Three subscales were derived from the DAS: achievement (14 items, \( \alpha = .92 \)), dependency (seven items, \( \alpha = .75 \)) and self-control/dependency (five items, \( \alpha = .58 \); Power et al., 1994).

**Procedure**
Screening packs mailed to referred clients consisted of the BDI, the IIP-C and a brief questionnaire relating to psychoactive medication and previous experience of psychotherapy or counselling. Clients who returned the packs and met the inclusion criteria were invited to attend an assessment interview.

The intake assessment interview was carried out a mean of four weeks after
screening information was received at the clinic. The structured interview included the DSM-IV schedule for assessing Axis I disorders (American Psychiatric Association, 1994), used to establish diagnosis of MDE and to check that such a diagnosis was not caused by bereavement, and the Schedules for Clinical Assessment in Neuropsychiatry (Janka, Uestuen, & Sartorius, 1994), used to establish the absence of excluding symptomatology, such as mania, misuse of alcohol or drugs, or psychosis. At the intake assessment, the clients also completed the BDI, the IIP-C and the DAS. Clients were assigned randomly to therapists at their intake assessment. The randomization procedure did, however, take into account therapist availability. Once it had been established that the inclusion criterion of the presence of MDE was met, and no excluding factors applied, the client read, signed and returned the consent form and completed the first set of assessment questionnaires. A first therapy appointment usually took place within two weeks of assessment.

Prior to each session, clients completed the BDI. Of 674 sessions attended by the 58 clients, BDI data were missing for 14 sessions. Two weeks after the final session, of the 35 clients who completed therapy, 30 completed a post-treatment assessment, again including the BDI, the IIP-C and the DAS. The remaining five clients failed to attend for their post-therapy assessment. In addition, one client who did not complete therapy completed a post-therapy assessment.

Results

Table 1 presents scores for the 58 clients who began treatment on these measures at screening and intake assessment.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Screening</th>
<th>Intake assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>BDI (total score)</td>
<td>31.00</td>
<td>6.84</td>
</tr>
<tr>
<td>IIP-C</td>
<td>2.32</td>
<td>0.87</td>
</tr>
<tr>
<td>DAS (total score)</td>
<td>162.88</td>
<td>34.79</td>
</tr>
<tr>
<td>DAS achievement</td>
<td>56.32</td>
<td>16.94</td>
</tr>
<tr>
<td>DAS dependency</td>
<td>34.86</td>
<td>6.78</td>
</tr>
<tr>
<td>DAS self control/dependency</td>
<td>17.18</td>
<td>5.19</td>
</tr>
</tbody>
</table>

Note. BDI = Beck Depression Inventory; IIP-C = Inventory of Interpersonal Problems, Cluster C Scale; DAS = Dysfunctional Attitudes Scale.

We tested for pre-treatment differences between the 12-session cohort and the 12-20 session cohort. No significant differences between the 12-session cohort and the 12-20 session cohort were found on any of the intake measures, nor did the two cohorts differ significantly on any of the demographic measures.

Statistically significant change

Clients who completed both therapy and the post-therapy assessment (N = 30),
achieved statistically significant improvement on the BDI as revealed by paired t tests examining mean differences between intake and post-therapy, $M = 17.26$, $t(29) = 7.28$, $p < .0001$, and on the IIP-C, $M = .80$, $t(27) = 5.48$, $p < .0001$. The mean scores are shown in Table 2.

Paired t tests examining the difference between intake assessment and final session BDI scores indicated that the average improvement across all clients who had received therapy was statistically significant, $M = 13.8$ BDI points, $t(56) = 8.90$, $p < .0001$; effect size (ES) = 1.27. Significant gains were made by both completers ($n = 35$), $M = 16.9$, $t(34) = 7.67$, $p < .0001$; ES = 1.71, and non-completers ($n = 22$), $M = 8.91$, $t(21) = 5.88$, $p < .0001$; ES = .89, as indicated by paired t tests examining mean differences between pre- and final session BDI. (One non-completer failed to take the BDI at the last session and was omitted from this analysis.)

A repeated-measures ANOVA to test for differences between completers and non-completers on the amount of change between intake and final-session BDI scores indicated significant advantages to the completers over non-completers, as shown by the interaction between status (completer or non-completer) and time (intake to final session), $F(1,55) = 6.90$, $p = .01$. The main effect of treatment (i.e. the change between intake and final session BDI scores) was highly significant, $F(1,55) = 72.50$, $p < .0001$, consistent with the separate t tests on completer and non-completer groups.

To provide a control for differences in number of sessions received, we matched each of the 22 non-completers’ final BDI score with a randomly chosen completer’s score at the corresponding session and used that as the dependent variable (e.g. if a non-completer dropped out at Session 4, we used the BDI of a completer at Session 4). With the number of sessions thus controlled, a repeated-measures ANOVA indicated that non-completers were not significantly different from completers on the amount of change achieved between intake and matched final session, as shown by the interaction between status (completer or non-completer) and time (intake to matched final session), $F(1,42) = .03$, $p = .87$, in part reflecting the fact that the non-completers’ BDI mean score was slightly (but non-significantly) higher at intake. Table 3 presents the BDI means for completers and non-completers at intake, last session and matched session.

**Table 2.** Mean scores pre- and post-intervention for clients who completed therapy

<table>
<thead>
<tr>
<th>Measure</th>
<th>Intake assessment</th>
<th>Post therapy assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>BDI</td>
<td>27.57</td>
<td>11.32</td>
</tr>
<tr>
<td>IIP-C</td>
<td>2.19</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Note. BDI = Beck Depression Inventory; IIP-C = Inventory of Interpersonal Problems, Cluster C Scale.

**Reliable and clinically significant change**

According to Jacobson and Truax (1991), conceptualizing clinically significant improvement yields two criteria. First, there must be **reliable change**, that is it is important to know how much change has occurred as a result of therapy over and above the fluctuations of an imprecise measuring instrument (see Jacobson & Truax, 1991, p. 14 for formulae). Second, there must be a **clinically significant result**, one
suggested definition of which was the client belonging to the functional or normal population—within two standard deviations of the normal mean—at the end of treatment. Thus, a client who begins treatment outside the normal limits and moves to some point within them during or following treatment, having changed by at least the reliable change index, can be said to have achieved reliable and clinically significant change.

To set an upper limit on the BDI total score for membership in the functional population, we used normative data published by Nietzel, Russel, Hemmings, and Gretter (1987), who gave a mean of 4.54 and standard deviation of 4.46 for a non-distressed population. Thus, we set the threshold for a clinically significant result two SDs above the non-distressed mean (i.e., a BDI score of 13.46). The reliable change threshold was 6.18—calculated as 1.96 times the standard error of the difference (estimated from the SD for the non-distressed population and the BDI’s average test-retest reliability of .75, from A. T. Beck, Steer, and Garbin, 1988). Thus, clients were held to have shown clinically significant change on the BDI if their intake score was 14 or higher (cf. inclusion criterion of BDI ≥ 18) and their final session score (a) had changed by at least 7 points, and (b) had dropped to 13 or below.

Table 3. BDI scores at intake, final session and matched session for completers and non-completers

| Assessment point | Completers | | | | | Non-completers | | | |
|------------------|------------|------------------|------------------|------------------|
|                  | M          | SD               | N                | M               | SD               | N                |
| Intake BDI       | 27.31      | 10.88            | 35               | 30.00           | 9.26             | 23               |
| Final BDI        | 10.46      | 10.01            | 35               | 21.91           | 9.24             | 22               |
| Matched BDI      | 15.91      | 10.66            | 22               | 21.91           | 9.24             | 22               |

Note. BDI = Beck Depression Inventory.

Figure 1. Plot of reliable and clinically significant change on the BDI.
Figure 1 presents an outcomes matrix (i.e. a scatter plot) showing the relation between intake BDI and final session BDI for completers (triangles) and non-completers (circles). The diagonal solid line represents no change. The parallel diagonal dotted lines represent the reliable change index limits (7 BDI points). Changes made by clients falling between these lines were too small to be considered as reliable. The area below the horizontal dotted line represents a clinically significant result (BDI 13 or below). As Fig.1 shows, a far higher proportion of completers (25/35; 71.4% of subgroup) than non-completers (3/23; 13% of subgroup) met these criteria for reliable and clinically significant change. A statistical comparison of these proportions was highly significant, \( \chi^2(1) = 18.96, p < .0001 \). Nevertheless, a large majority of the non-completers (16/23; 69.6%) did experience reliable degrees of improvement (shown by circles below the lower diagonal dotted line).

### Discussion

Our overall results suggest that CT is an effective psychological intervention for clients presenting with depression in a routine clinic setting. They also suggest that clients who complete their agreed number of sessions improve more than clients who drop out of treatment.

The pre-post ES of 1.27 overall and 1.71 for completers was somewhat less than pre-post ESs obtained from previous RCTs comparing versions of CT with alternative therapies. For example, the pre-post ES for the BDI in the NIMH TDCRP (Elkin, 1994) for CT was 1.98 (intake \( M = 26.8, SD = 8.4 \); post-treatment \( M = 10.2, SD = 8.7 \)). This analysis was based on an intent-to-treat sample, so the more appropriate ES for comparison is that derived from our overall ES (i.e. 1.27). In a UK comparative trial, the pre-post ES for the BDI in the Second Sheffield Psychotherapy Project (Shapiro et al., 1994) for 16 sessions of CT was 2.44 (intake \( M = 29.60, SD = 7.38 \); post-treatment \( M = 11.58, SD = 6.76 \)). This study used only clients who had completed treatment, so the ES for our completer sample is the more appropriate comparator (i.e. 1.71). Thus, the ESs from our study were lower than in these RCTs by about .70 in both comparisons.

However, in a replication of the Second Sheffield Psychotherapy Project carried out in NHS settings, the pre-post ESs for 16 sessions of CT was 1.66, an effect size broadly similar to that obtained in the present study for completers (Barkham, Rees, Shapiro et al., 1996). These results converge with ours and with those of Hirsch, Jolley, and Williams (2000) in suggesting that CT, while still very effective, may be less effective, on average, under conditions that resemble routine practice than under conditions in highly controlled efficacy trials. However, Shadish et al. (2000) warn against failing to discriminate between two research agendas: **effectiveness** (i.e. acquiring empirical support for the effectiveness of psychological therapies under clinically representative conditions) and **efficacy** (i.e. building evidence for empirically supported psychological therapies). The latter refers to clearly specified psychological therapies which have been proven to be efficacious in a research setting. These two contexts comprise different activities and should be evaluated as such, a fact which would account to some degree for poorer results for clinically representative therapies on a continuum of effect sizes.
Completers versus non-completers

The finding that completers obtained overall better gains—that is, a better end-state—than non-completers is consistent with commonsense expectations and clinical lore. We found that 71% of the completers, but only 13% of the non-completers, achieved reliable and clinically significant change on their final session BDI, suggesting that they had changed enough across therapy to rejoin the normal population in this respect. We note, however, that when completers and non-completers were compared on the amount of statistically significant change achieved after matched numbers of sessions (p. 12), there was no significant difference between the two groups. These results thus seem broadly consistent with the expectation that the effect of therapy reflects the dose of therapy, measured as the number of sessions (Barkham, Rees, Stiles et al., 1996; Howard, Kopta, Krause, & Orlinsky, 1986; Shadish et al., 2000). However, it should be noted that analyses based on statistical change (as opposed to clinical and reliable change) and on group averages (as opposed to individual plots)—(see Figure 1) are less stringent.

Even among clients who did not complete treatment, most (70%) achieved reliable degrees of improvement (see Fig. 1). Perhaps for some clients, this type of improvement was sufficient, and they chose to leave treatment. However, unresolved or sub-threshold symptoms are likely to make clients more vulnerable to a subsequent episode of depression (Wells, Burnham, Rogers, Hays, & Camp, 1992). Such recurrences of depression increase the burden of disease already carried by society through health care costs and lost economic productivity (Murray & Lopez, 1997).

Finally, we make comment on two methodological points of the study. First, we note that the comparison between completers and non-completers was made possible by repeated administration of the BDI, which made it possible to assess the improvement achieved by non-completers at the point of their last attended session. To the extent that they overlook the outcomes of non-completers, traditional pre–post assessment designs may provide misleadingly positive results, in that they are likely to be based on the sub-population of well-motivated and committed clients who complete therapy. More generally, the issue of how to ensure capture of a high percentage of post-therapy data in routine practice is a considerable challenge. Second, the clinical relevance of the different outcomes between completers and non-completers was made clear by the use of methods to determine reliable and clinically significant change. This procedure is consistent with Kendall’s (1997) statement that ‘Evaluations of the outcomes of psychological treatments are favourably enhanced when the published report includes not only statistical significance and the required effect size but also a consideration of clinical significance’ (p. 3). Moreover, presenting these data in graphical form (recall Fig. 1) allows implicit evidence of the confidence by which any one of the particular criteria was met. Still further, it preserves both the continuous nature of the phenomenon (as represented by the BDI) and the individuality of the clients (each point represents an individual).

The value of adopting the criteria of reliable and clinically significant change and presenting it in graphical form is borne out by the way in which completers and non-completers were so clearly distinguished in this study. The magnitude of the differences between the two groups of clients was made clear by capturing the outcomes of each individual client. In contrast, standard inferential statistics do not provide full information on differences between comparison groups, based as they are on group averages. Furthermore, while paired t tests demonstrated that non-completers as a group made a statistically significant improvement, Fig. 1 differentiated non-completers
according to the criteria of reliable and clinically significant change. Addressing these two methodological points (i.e. how to capture post-therapy data, and how to present findings in clinically sensitive and meaningful ways) will go some way towards progressing the efficacy-effectiveness dilemma current within the clinical and psychotherapy literature (e.g. Nathan, Stuart, & Dolan, 2000) and do much to inform the ‘practice-based evidence’ of CT for depression.

References


Received 9 February 2001; revised version received 1 May 2002