Does cognitive-behavioural therapy influence the long-term outcome of generalized anxiety disorder? An 8–14 year follow-up of two clinical trials

R. C. DURHAM,[1] J. A. CHAMBERS, R. R. MacDONALD,
K. G. POWER AND K. MAJOR

*From the Department of Psychiatry, University of Dundee and Department of Clinical Psychology,
Tayside NHS, Dundee; Department of Psychology, University of Stirling; and Department of Health Policy,
Ayrshire and Arran NHS, Ayr*

ABSTRACT

Background. Generalized anxiety disorder is a common condition of excessive worry and tension which tends to run a chronic course associated with significant psychiatric and medical problems. Cognitive behaviour therapy (CBT) has been shown to be of clinical value in about 50 % of cases with treatment gains maintained over follow-up periods ranging from 6 to 12 months. The potential value of CBT over the longer term has not been subject to rigorous investigation.

Method. Results are reported of 8–14 year follow-up of two randomized controlled trials of cognitive-behaviour therapy for generalized anxiety disorder employing structured interview with an assessor blind to initial treatment condition. Comparison groups included medication and placebo in one study based in primary care, and analytical psychotherapy in the other based in secondary care. Follow-up samples (30 % and 55 % of trial entrants) were broadly representative of the original cohorts.

Results. Overall, 50 % of participants were markedly improved of whom 30–40 % were recovered (i.e. free of symptoms). Outcome was significantly worse for the study based in secondary care in which the clinical presentation of participants was more complex and severe. For a minority (30–40 %), mainly from the secondary care study, outcome was poor. Treatment with CBT was associated with significantly lower overall severity of symptomatology and less interim treatment, in comparison with non-CBT conditions, but there was no evidence that CBT influenced diagnostic status, probability of recovery or patient perceptions of overall improvement.

Conclusions. Both CBT and the complexity and severity of presenting problems appear to influence the long-term outcome of GAD.

INTRODUCTION

Generalized anxiety disorder (GAD) is a prevalent condition of excessive worry and tension which typically runs a fluctuating and often chronic course associated with heavy demands on health service resources (Ballenger *et al.* 2001). The duration and severity of particular episodes are influenced by a range of biological and psychosocial variables including psychological therapy (Hoehn-Saric *et al.* 1995). Fisher & Durham (1999) analysed the clinical significance of changes following psychological therapy for GAD, using Jacobson criteria, and found that overall rates of recovery at 6 month follow-up were relatively modest (40 %) although highest with individual cognitive-behavioural therapy (CBT) and applied relaxation (50–60 %). Borkovec & Ruscio (2001) reviewed 13 controlled

[1] Address for correspondence: Dr R. C. Durham, Department of Psychiatry, University of Dundee, Ninewells Hospital & Medical School, Dundee DD1 9SY.

clinical trials of psychological therapy for GAD and found that CBT was associated with the lowest dropout rates and the largest within-group and between-group effect sizes relative to all other comparison conditions. They found consistent evidence that therapeutic gains following CBT were maintained but length of follow-up averaged only 9 months with a range of 6 to 12 months.

It is worth noting that all of the studies in these reviews were conducted by clinical researchers with a primarily cognitive-behavioural orientation; a less promising picture is reported by Seivewright and colleagues who conducted a 5-year follow-up of a cohort of 210 psychiatric out-patients suffering from GAD, panic disorder or dysthymic disorder and randomized to medication, CBT or self-help (Seivewright et al. 1998). Sixty per cent had a broadly favourable outcome with the remainder handicapped either intermittently or continuously throughout the follow-up period. Initial diagnosis and original treatment were found to be of no predictive value although a further analysis found that over half the patients had been treated by therapists of uncertain competence and that lasting benefit was associated with CBT delivered by therapists rated as competent during supervision (Kingdon et al. 1996).

Clearly CBT, with its emphasis on teaching active coping skills, has the potential to influence the long-term course of GAD. Current evidence, however, consists mainly of postal questionnaires of symptom severity over relatively short follow-up periods, with variable definitions of outcome, and is too limited to draw firm conclusions (Roth & Fonagy, 1996; Tyrer, 2000). Moreover, studies of the natural history of anxiety disorders tend to suggest that the long-term outcome of GAD is generally less good than other types of anxiety disorder with less than one-third achieving recovery (Ormel et al. 1993; Quinton et al. 1995). A true picture of the power of psychological therapies to alter specific conditions will only emerge from comprehensive assessments of representative samples of patients randomized to different treatment conditions and followed up over a number of years.

In this paper we report results from an extended follow-up of two clinical trials of CBT for GAD (Power et al. 1990; Durham et al. 1994) in which self-report data were supplemented by a structured interview with an assessor blind to initial treatment condition. Two broad questions are addressed: (1) What is the long-term outcome of patients who participated in these two clinical trials?; and (2) Does treatment with CBT confer an advantage relative to medication and alternative psychotherapy?

METHOD

Participants

Study 1 participants took part in a randomized, controlled trial of diazepam (DZ), placebo (PL), cognitive behavioural therapy alone (CBT), cognitive behavioural therapy plus diazepam (CBT+DZ) and cognitive behavioural therapy plus placebo (CBT+PL). The trial was based at the Anxiety and Stress Centre, University of Stirling, between 1985 and 1988. All patients were recruited from primary care health centres and had a primary diagnosis of generalized anxiety disorder (GAD) according to DSM-III criteria (American Psychiatric Association, 1980). A full description of the trial can be found in the original report (Power et al. 1990). Study 2 participants took part in a randomized, controlled trial of cognitive therapy (CT), analytical psychotherapy (AP) or anxiety management training (AMT). The trial was based at the Department of Psychiatry, University of Dundee between 1989 and 1991. Patients were recruited from either general physicians in primary care or from psychiatrists in secondary care and had a primary diagnosis of GAD according to DSM-III-R criteria (American Psychiatric Association, 1987). A full description of the trial can be found in the original report (Durham et al. 1994). In neither study was co-morbidity controlled for in allocating cases to treatment groups.

In both studies there was an attempt to follow-up all participants since it was felt that drop-outs, non-attenders and completers were equally important in gaining a comprehensive view of the long-term course of GAD on an intention-to-treat basis. Of the 111 patients entering Study 1, 10 had failed to attend for treatment, 18 had dropped out of treatment before the mid-point and 83 had completed treatment. Seventeen of these 111 patients were either untraceable at long-term follow-up or had died in the interim period, and one GP refused permission to contact a patient. Of the 93 patients who were

contacted, 33 (30 %) took part in the long-term follow-up study, with 28 of these attending for full diagnostic interview. Of the 110 patients entering Study 2, 11 had failed to attend for treatment, 19 had dropped out of treatment and 80 had completed treatment. Sixteen of these 110 patients were either untraceable at long-term follow-up or had died in the interim period, and one GP refused permission to contact a patient. Of the 93 patients who were contacted, 61 (55 %) took part in the long-term follow-up study, with 51 of these attending for full diagnostic interview.

Participant characteristics

The follow-up sample in Study 1 consisted of 14 men and 19 women with a mean age of 53 years (range 30 to 72) who were interviewed between 11 and 14 years following discharge from treatment in the original trial. The sample in Study 2 consisted of 20 men and 41 women with a mean age of 48 years (range 27 to 71) who were interviewed between 8 and 10 years following discharge from treatment in the original trial. Study 2 patients had significantly higher scores on a seven-point measure of degree of social deprivation based on postcode sector (Carstairs & Morris, 1991) (mean difference = 1·2, 95 % CI (0·5, 1·9), $P < 0·01$) and had a greater chronicity of disorder as might be expected from the longer duration criterion used in DSM-III-R in comparison with DSM-III (6 months *versus* 1 month).

Comparison of participants and non-participants

In view of the large number of patients who were unavailable to follow-up, or who failed to respond to a postal invitation to take part, an exhaustive analysis was made of the characteristics of participants and non-participants in respect of baseline demographics, allocation to treatment conditions, baseline symptomatic state and response to treatment. Full details of these comparisons are available from the senior author. In summary, there were no significant differences between follow-up participants and non-participants with regard to age, gender, marital status, social deprivation, frequency of visits to GPs (primary care physicians) before and after the original trials, presence of co-morbid diagnoses at the time of the original trial, family history of mental health problems, original measures of symptom severity, proportions

in original treatment conditions, proportions on medication for anxiety at the time of the original trials or proportions with previous psychological or psychiatric treatment prior to the original trials.

Participants from Study 2, however, were significantly more likely than non-participants to have had completed initial treatment (84 % *v.* 59 %) ($\chi^2(2) = 8·4$, $P < 0·05$). Participants from Study 1 were significantly more likely to have had a positive response to previous drug treatment for anxiety disorders than non-participants (88 % *v.* 59 %) ($\chi^2(1) = 5·8$, $P < 0·05$) and to have had a shorter duration of current episode of anxiety at the time of the trial (either < 1 month or between 1 and 3 months) than non-participants (73·1 % *v.* 41·7 %; $\chi^2(3) = 9·1$, $P < 0·05$).

Procedure

The same overall procedure was followed in both studies as approved by the respective medical research ethics committees in the two localities. Patient details from the original trials were verified with local medical records and attempts were made to trace patients who had moved using central medical records where available. Permission to contact all successfully traced patients was then sought from each patient's general medical practitioner and if this was refused ($N = 2$) the patient was excluded. Patients residing in Scotland were contacted by letter and invited to complete the main outcome questionnaires and to participate in a one-hour interview if they were in reasonable travelling distance. One patient living outside Scotland was sent questionnaires only. Every effort was made to see participants at a time and place of their convenience either at the research base, a local clinic or the patient's home.

Interviews were all conducted by the same research psychologist (J.C.) who was blind to treatment condition until the very end of the interview when there was a brief assessment of attitudes to the original treatment to be reported in a separate paper. The interview consisted of: (1) the Anxiety Disorders Interview Schedule IV (ADIS-IV; Brown *et al.* 1994) to assess diagnostic status according to DSM-IV criteria; and (2) an assessment of amount of treatment for mental health problems since participation in the original trial. The following four-point scale was used to categorize the amount of

patient-reported interim treatment for anxiety with either medication or psychological therapy: (1) none; (2) little – treatment over one short time period; (3) moderate – treatment over several years; and (4) a lot – treatment for the majority of the intervening period. This scale has some similarities to that used by Seivewright and colleagues (1998) to categorize overall outcome on the basis of case note review but, unlike their scale, excludes symptom severity and social dysfunction. Many patients had difficulty recalling the precise nature of treatment received and it was felt that reliable data could only be collected for the overall amount. Patient responses on amount of interim treatment were cross-checked with medical case notes for 36% (N = 12) of the sample in Study 1. In eight cases (67%) results tallied exactly, and in the remaining four cases three had underestimated the amount of treatment received by one category, and one had underestimated it by two categories.

Those patients who failed to respond to the initial letter within 3 weeks were sent a reminder, which included the main outcome questionnaires. This letter offered those patients not wishing to attend an interview the additional options of either a telephone interview plus questionnaires or the completion of questionnaires only.

Measures

Outcome measures administered to all participants were as follows.

Clinical Global Severity (CGS)

This is a component of the ADIS-IV and was measured on 0–8 point scale (0 = no evidence of disorder, 4 = definitely disturbing/disabling, and 8 = very severely disturbing/disabling).

Brief Symptom Inventory (BSI) (Derogatis, 1993)

This is a shortened version of the SCL-90 and a global measure of current symptomatic state, measured on a five-point scale (0–4). The BSI global severity index (BSI-GSI), representing a mean score for all completed items, was used in the current analysis. Jacobson criterion (c) (Jacobson & Truax, 1991) was used to identify clinically significant change on this measure defined as a reliable change between pre- and follow-up scores and a score below the cut-off

point for normative functioning. For criterion (c), the cut-off is calculated as the mid-point between two standard deviations below the pre-treatment mean and two standard deviations above the mean for the normal population.

Clinical Global Improvement (CGI)

This is a patient-rated seven-point scale measuring change in condition since the original trial using the following categories: marked/moderate/mild improvement/no change/mild/moderate/marked deterioration (Guy, 1982).

UK SF-36 version 2 (Jenkinson et al. 1999)

The SF-36 Health Survey was developed by Ware and colleagues (1993) as a measure of health status comprising eight subscales: physical role, physical functioning, emotional role, vitality, bodily pain, mental health, social role and general health. Jenkinson et al. (1999) developed norms for a UK population, using a version of the SF-36, which has minor wording differences from the original measure, in order to make it more acceptable to UK populations. The physical (SF36-PC) and mental (SF36-MC) components of the SF-36, originally derived via factor analysis, summarize the eight original subscales. The SF36-MC was used in the current analysis.

In addition the clinician-rated Hamilton Rating Scale for Anxiety (HAM-A) (Hamilton, 1959) was included for Study 1 only, using the structured scoring system used in the original trial (cf. Power *et al.* 1985).

Because there was no overlap in self-report questionnaires used in the original two studies a main outcome measure specific to each study was also included as follows: the Kellner & Sheffield (1973) Symptom Rating Test (SRT), a questionnaire designed to measure changes in symptoms of distress in neurotic patients, was used in Study 1 but not Study 2; and the trait version of the State Trait Anxiety Inventory (STAI-T) (Spielberger *et al.* 1983) was used in Study 2 but not Study 1.

RESULTS

Changes in symptom severity from pre-treatment to follow-up on original outcome measures

Table 1 shows the means and standard deviations of the main outcome measures before treatment

Table 1.   *Changes in symptom severity from pre-treatment to long-term follow-up on original outcome measures for each study*

Treatment group	N	Pre-treatment Mean (s.d.)	Long-term follow-up Mean (s.d.)	Difference	(95% CI)
Study 1					
HAM-A†					
CBT alone	7	14·7 (3·4)	7·9 (3·1)		
CBT + DZ	8	12·9 (5·8)	9·1 (6·5)		
CBT + PL	4	16·0 (4·1)	9·0 (6·7)		
CBT group	19	14·2 (4·6)	8·6 (5·2)	− 5·6***	− 7·7, − 3·5
DZ only	7	13·4 (6·4)	11·4 (5·6)		
PL only	3	16·7 (2·1)	10·3 (7·5)		
Non-CBT	10	14·4 (5·6)	11·1 (5·8)	− 3·3	− 9·1, 2·5
SRT†					
CBT alone	8	35·4 (18·3)	18·9 (13·3)		
CBT + DZ	9	29·2 (16·8)	19·8 (13·0)		
CBT + PL	4	32·8 (12·5)	18·5 (16·3)		
CBT group	21	32·2 (16·2)	19·2 (13·0)	− 13·0**	− 20·2, − 5·9
DZ only	7	30·1 (14·0)	23·0 (17·5)		
PL only	4	38·0 (9·3)	29·8 (17·8)		
Non-CBT	11	33·0 (12·6)	25·4 (17·0)	− 7·6	− 16·9, 1·8
Study 2					
BSI-GSI†					
CT	22	1·73 (0·7)	1·15 (0·8)		
AMT	11	1·45 (0·5)	1·16 (0·8)		
CBT group	33	1·64 (0·6)	1·15 (0·8)	− 0·48**	− 0·71, − 0·25
AP	23	1·74 (0·6)	1·37 (0·8)		
Non-CBT	23	1·74 (0·6)	1·37 (0·8)	− 0·38**	− 0·73, − 0·02
STAI-T†					
CT	22	58·6 (9·9)	49·8 (14·1)		
AMT	12	52·8 (14·0)	48·5 (12·9)		
CBT group	34	56·6 (11·6)	49·4 (13·5)	− 7·2**	− 11·1, − 3·3
AP	23	61·1 (6·2)	49·2 (13·9)		
Non-CBT	23	61·1 (6·2)	49·2 (13·9)	− 11·9***	− 16·7, − 7·1

† Lower scores are more favourable.
Paired *t* tests: ** $P < 0.01$; *** $P < 0.001$.

and at long-term follow-up for each treatment group. Only those patients for whom scores were available at both time points are included in Table 1. For the purposes of analyses the treatment groups were amalgamated into CBT and non-CBT for each study as follows: Study 1 – CBT, CBT + DZ, CBT + PL *v.* DZ and PL; Study 2 – CT and AMT *v.* AP. The groups were amalgamated because the number of patients in each treatment group was small and separate one-way ANOVAs on the differences between the individual treatment groups did not reveal any significant differences on any follow-up outcome measure in either study. Table 1 shows that there was a reduction in symptom severity in all treatment groups in both studies. Paired *t* tests on the change from pre-treatment to long-term follow-up show a significant improvement in the combined CBT group in Study 1 on both the HAM-A and the SRT, but no significant improvement on either measure for the non-CBT group. The same analysis showed a significant improvement for both the CBT and the non-CBT groups in Study 2 on both the BSI-GSI and the STAI-T. Although in general the CBT groups appeared to improve more than the non-CBT groups, independent *t* tests revealed these differences to be non-significant in both studies.

Proportions sustaining and achieving recovery from 6 month to long-term follow-up

Table 2 shows the proportion of long-term follow-up participants who achieved or maintained recovery according to Jacobson criteria on the main self-report outcome measure for each study, broken down by CBT group. Jacobson defined clinically significant change as achieving a score below a cut-off point representative of normative functioning, together with a reliable degree of change between pre-treatment and follow-up scores (Jacobson & Truax, 1991). In both studies we repeated the analysis using the

Table 2. *Recovery rates at 6-month and long-term follow-up using Jacobson criteria applied to the primary self-report outcome measure in each study*

| Treatment group | 6 mth follow-up | | Long-term follow-up | | | | | |
| | Recovered | | Recovery maintained | | Recovery achieved | | Recovery rate overall | |
	N	(%)*	N	(%)*	N	(%)*	N	(%)*
Study 1: SRT†								
CBT	16	(80)	7	(35)	1	(5)	8	(40)
Non-CBT	7	(70)	3	(30)	0	(0)	3	(30)
Study 2: STAI-T‡								
CBT	8	(26)	3	(10)	8	(26)	11	(36)
Non-CBT	2	(11)	2	(11)	6	(32)	8	(42)

* The number achieving or maintaining recovery as a percentage of long-term follow-up participants in each group as follows: Study 1 CBT = 20, non-CBT = 10; Study 2 CBT = 31, non-CBT = 19.
† SRT, Kellner & Sheffield Symptom Rating Test.
‡ STAI-T, trait version of the State Trait Anxiety Inventory.

Jacobson criteria adopted in the original trials. For study 1, a cut-off point of 15 on the SRT was used, which was equivalent to two standard deviations below the sample pre-treatment mean. In Study 2, we used Jacobson criterion (c) which gave a cut-off score of 47 on the STAI-T, together with a reliable change index equating to a change from pre-treatment to long-term follow-up of $\geqslant 8$.

Table 2 shows the actual numbers of long-term follow-up participants achieving or maintaining recovery, also shown as a percentage of the total number of participants in each group. It can be seen in Study 1, that while the majority of patients (70–80%) had achieved clinically significant change on the SRT at 6 month follow-up, less than half of these (i.e. 10 out of 23) had maintained their recovery at long-term follow-up. In addition, of those not recovered at 6 month follow-up, only one patient had achieved recovery status in the interim period, and the overall recovery rate at long-term follow-up was therefore modest (30–40%).

In Study 2, less than a quarter of long-term follow-up participants had achieved recovery at 6 month follow-up according to Jacobson criteria on the STAI-T, and only half of these (i.e. 5 out of 10) had maintained that recovery to long-term follow-up. In Study 2, 14 patients who were not recovered on the STAI-T at 6 month follow-up did achieve recovery at long-term follow-up, giving overall recovery rates at long-term follow-up of 36–42%.

Finally, it can be seen from Table 2 that there are no appreciable differences in recovery rates between the CBT and non-CBT groups for either study.

Comparison of Study 1 and Study 2

Although nearly one half of the total sample received no DSM-IV diagnosis the proportion of Study 1 patients with no diagnosis was nearly double that of Study 2 (69% v. 37%) and this difference was significant. Of those patients who did meet diagnostic criteria for one or more disorders the majority were co-morbid (79%) with co-morbidity being much more common in Study 2 than in Study 1. The nature of the diagnoses is of some interest. Only one individual in the total sample had a pure diagnosis of GAD and there is evidence of drift toward non-GAD diagnoses with 10% of the total sample meeting criteria for a single non-GAD diagnosis (social phobia = 3, panic disorder = 3, hypochondriasis = 1, depression = 1, simple phobia = 2) and 8% meeting criteria for more than one non-GAD diagnosis.

Nearly two-thirds of the total sample reported receiving little or no interim treatment for mental health problems since the original trials with 20% reporting a moderate amount and 16% reporting a lot. Again, there is a significant difference in the pattern of interim treatment across the two studies with a sizeable minority of Study 2 patients receiving a lot of treatment (23%) and a negligible proportion so doing in Study 1 (3%) ($\chi^2(1) = 6.3$, $P < 0.05$).

Across both studies, the proportion of patients defined as 'recovered' or 'in remission' on the Brief Symptom Inventory (BSI) was measured

Table 3. *Patient status at long-term follow-up by study and by amalgamated CBT and non-CBT groups*

| | Study 1 | | Study 2 | | Study 1 v. Study 2 | | Total sample | | | | | | CBT v. Non-CBT | |
| | | | | | | | CBT | | Non-CBT | | All | | | |
	N	(%)	N	(%)	df	χ²	N	(%)	N	(%)	N	(%)	df	χ²
Diagnostic status														
None	20	(69)	20	(37)			24	(48)	16	(48)	40	(48)		
One diagnosis	4	(14)	5	(9)			5	(10)	4	(12)	9	(11)		
Two or more diagnoses	5	(17)	29	(54)	2	10·5**	21	(42)	13	(40)	34	(41)	2	0·1
Interim treatment														
None/little	23	(70)	37	(61)			38	(68)	22	(58)	60	(64)		
Moderate	9	(27)	10	(16)			13	(23)	6	(16)	19	(20)		
A lot	1	(3)	14	(23)	2	6·9*	5	(9)	10	(26)	15	(16)	2	5·3
Clinically significant change on BSI-GSI														
BSI non-case†	11	(33)	13	(22)	1	1·4	18	(33)	6	(16)	24	(26)	1	3·6
BSI Jacobson criteria‡	16	(48)	23	(39)	1	0·8	26	(48)	13	(34)	39	(42)	1	1·8
Patient ratings of improvement														
Marked improvement	20	(69)	23	(43)			25	(50)	18	(55)	43	(52)		
Moderate/mild improvement	8	(28)	15	(28)			16	(32)	7	(21)	23	(28)		
No change/worse	1	(3)	16	(30)	2	8·8*	9	(18)	8	(24)	17	(20)	2	1·3

† BSI caseness as defined by Derogatis (1993).
‡ Jacobson criterion (c) as defined for Study 2 used across both studies (without reliable change index).
* $P < 0.05$; ** $P < 0.01$.

in terms of both the percentage of participants achieving caseness as defined by Derogatis (1993) and the percentage of participants below the cut-off (0·98) for clinically significant change as defined by Jacobson criterion (c) as calculated in Study 2 (Jacobson & Truax, 1991). The reliable change index was included for Study 2 only (equating to a change from pre-treatment to long-term follow-up of $\geqslant 0.51$) since pre-treatment scores on the BSI were unavailable for Study 1. It can be seen that the majority of participants remained symptomatic in varying degrees. Although higher proportions of Study 1 than Study 2 patients achieved recovery status the differences were not significant.

Table 3 also shows that there were significant differences in the patient ratings of improvement between the studies. About one half of the total sample described themselves as markedly improved, with 69% in Study 1 compared with 43% in Study 2. This difference is significant $(\chi^2(1) = 5.3, P < 0.05)$.

Comparison of CBT and non-CBT groups

There were no significant differences with respect to diagnostic status. There was a tendency for patients in the non-CBT groups to have had a greater amount of interim treatment, and,

although the overall χ^2 was not significant, only 9% of the CBT group reported having a lot of treatment *versus* 26% of the non-CBT group, and this difference was significant $(\chi^2(1) = 5.1, P < 0.05)$. A higher percentage of those in the CBT group were defined as 'non-cases' (Derogatis, 1993) on the BSI (33% v. 16%) and this difference approached significance $(P = 0.059)$ (see Table 3). In addition a higher percentage of participants from the CBT group had achieved the cut-off score on the BSI as defined by Jacobson criterion (c) for Study 2 (48% v. 34%), but this difference was not significant. There were no difference between the CBT and non-CBT groups in patient perceptions of overall improvement.

Cross-sectional analysis of outcome at long-term follow-up on common measures of overall symptom severity

The mean scores of the main outcome measures of symptom severity at long-term follow-up (i.e. the CGS, BSI-GSI and SF36-MC) are shown in Table 4. These are broken down by amalgamated CBT and non-CBT groups for each study. Table 4 also shows the pre- and post-treatment scores on the primary outcome measure used in each of the original trials (i.e. the HAM-A for Study 1

Table 4. *Observed and estimated means and standard errors of the means for symptom severity measures at long-term follow-up by amalgamated CBT and non-CBT groups for Study 1 and Study 2*

| | Study 1 | | | | | | Study 2 | | | | | |
| | Observed data | | | | Direct ML estimates | | Observed data | | | | Direct ML estimates | |
	N	CBT Mean (s.e.)	N	Non-CBT Mean (s.e.)	CBT Mean (s.e.)	Non-CBT Mean (s.e.)	N	CBT Mean (s.e.)	N	Non-CBT Mean (s.e.)	CBT Mean (s.e.)	Non-CBT Mean (s.e.)
HAMA pre*	61	18·48 (0·49)	42	18·38 (0·49)	18·48 (0·49)	18·38 (0·49)						
HAMA post*	57	4·77 (0·58)	23	7·52 (0·82)	4·75 (0·58)	7·59 (0·73)						
BSI-GSI pre*							61	1·62 (0·08)	61	1·71 (0·10)	1·62 (0·08)	1·73 (0·09)
BSI-GSI post*							47	1·01 (0·12)	28	1·60 (0·16)	1·04 (0·12)	1·59 (0·15)
BSI-GSI*	21	0·82 (0·13)	12	1·43 (0·27)	0·91 (0·12)	1·48 (0·26)	33	1·15‡ (0·14)	26	1·53‡ (0·18)	1·18§ (0·13)	1·47§ (0·16)
CGS (0–8)*	19	1·95 (0·39)	10	2·60 (0·50)	2·40 (0·36)	3·25 (0·43)	31	3·58§ (0·35)	23	3·70‡ (0·42)	3·46§ (0·32)	3·73§ (0·44)
SF36-MC†	18	47·54 (2·59)	12	42·04 (3·38)	45·72 (2·53)	39·88 (2·96)	33	43·44‡ (1·94)	25	36·00‡ (2·48)	43·14§ (1·88)	35·64§ (2·52)

pre, Pre-treatment values; post, post-treatment values; all other values are at long-term follow-up.
* Lower scores are more favourable.
† Higher scores are more favourable.
‡ These CBT and non-CBT means differed significantly on a MANOVA ($F(3,47)=4·0$, $P=0·012$). Tested on its own the SF36-MC difference was significant ($F(1,49)=7·0$, $P=0·011$).
§ A combined test of the differences between these CBT and non-CBT means in AMOS yielded $\chi^2(3)=9·2$, $P=0·027$. Tested on its own the SF36-MC difference was significant ($\chi^2(1)=5·2$, $P=0·022$).

and the BSI-GSI for Study 2) together with the Direct Maximum Likelihood (DML) estimates of these statistics resulting from a missing data analysis described below.

It can be seen that participants in Study 1 had more favourable scores than participants in Study 2 on all three outcome measures at long-term follow-up. A MANOVA showed the differences between studies to be significant ($F(3,73)=4·1$, $P<0·05$) and significant univariate effects were found on the CGS ($F(1,75)=10·6$, $P<0·01$) and the SF36-MC ($F(1,75)=7·4$, $P<0·01$).

A MANOVA on the differences between the CBT and non-CBT groups combining the data from the two studies at long-term follow-up on the CGS, BSI-GSI and SF36-MC yielded a significant overall $F(3,73)=4·4$, $P<0·01$. Significant univariate effects were found on the SF36-MC ($F(1,75)=8·1$, $P<0·01$) and the BSI-GSI ($F(1,75)=5·8$, $P<0·05$).

Missing data analyses

As only 52% of the patients in the study participated in the long-term follow-up an attempt was made to determine what the data would have been if there was no missing data, by obtaining direct maximum likelihood (DML) estimates of the means of the CBT and non-CBT groups in each study using pre-, post- and follow-up scores. For each variable the method adjusts the means to take account of how missing patients scored on the remaining variables used in the analysis. It assumes that the data is random with respect to all other variables. The analysis is recommended in Allison (2001). We used the AMOS package to perform the calculations and to perform statistical tests on the differences between the CBT and non-CBT groups on the means of the long-term follow-up variables. The two studies had to be analysed separately because different measures were taken at the start of the studies. Owing to the limited amount of data available AMOS was unable to estimate the means simultaneously on all the available variables. It was therefore decided to perform the missing data analysis using the pre- and post-treatment scores for the primary outcome measure used in each original study (i.e. the HAM-A for Study 1 and the BSI-GSI for Study 2) together with the long-term follow-up outcome measures used across both studies (i.e. the BSI-GSI, CGS and SF-36-MC).

The means of all the observed data and the DML estimated means are given in Table 4. The most obvious difference between the two sets of means is that in Study 1 the DML means are worse than the observed means indicating that patients present at long-term follow-up were healthier than those absent. In both studies it can be seen that all the long-term follow-up means are better in the CBT group than in the non-CBT group in both the observed and DML means. Indeed, it appears that the missing data analyses have had little effect on the CBT/non-CBT differences. Table 4 shows that these differences were significant in Study 2 for both the observed and DML means. In Study 1 the overall tests of CBT/non-CBT differences were not significant at the 5% level, perhaps because Study 1 had more missing data, but there was a marginally significant difference on the BSI-GSI in the observed means ($F(1,24) = 3\cdot44$, $P = 0\cdot076$) and in the DML estimates ($\chi^2(1) = 3\cdot3$, $P = 0\cdot068$). It may be worth noting that when other long-term follow-up variables (i.e. the SRT and HAM-A in Study 1 and the STAI-T in Study 2) replaced those used in the missing data analyses the DML estimates consistently showed the CBT groups to do better than the non-CBT ones, although none of those variables was significant on a univariate test.

Interim treatment and long-term outcome

Overall, higher levels of interim treatment for anxiety were related to poorer outcome at long-term follow-up. Those patients with one or more clinical diagnoses at long-term follow-up were significantly more likely to have had a lot of interim treatment (5% with no diagnosis *versus* 28% with one or more diagnoses) ($\chi^2(1) = 7\cdot8$, $P < 0\cdot01$), and those with no clinical diagnosis at long-term follow-up were significantly more likely to have received no treatment for anxiety in the interim period (53% with no diagnosis *versus* 19% with one or more diagnoses) ($\chi^2(1) = 10\cdot5$, $P < 0\cdot01$).

DISCUSSION

It must be acknowledged at the outset that our relatively low overall contact rate, 55% of participants in Study 2 but only 30% of participants in Study 1, opens up the possibility that the follow-up samples may not be representative of the original cohorts (Flick, 1988). Exhaustive analysis revealed no evidence of bias in respect of a large range of possible variables but we did find a significant tendency for a higher proportion of Study 2 participants to have completed treatment and for a higher proportion of Study 1 participants to have had a positive response to previous drug treatment and a shorter duration of disorder at the time of the original trial. The results may, therefore, present a more positive picture of overall outcome than is actually the case, particularly in Study 1. In regard to the potential bias of 'investigator allegiance' (Luborsky *et al.* 1995) it should be noted that outcome assessments were made by a researcher who was blind to initial treatment condition.

Evidence of long-term clinical outcome was collected from several contrasting perspectives and the results were broadly consistent across measures with marked differences in outcome between the two studies. In Study 1 about 70% had no diagnosable clinical disorder, about 70% reported a marked improvement since the original trial and 33–48% were defined as 'recovered' on the main outcome measures. In Study 2, in contrast, only 37% had no diagnosable disorder, 43% reported a marked improvement since the original trial and 22–39% were defined as 'recovered' on the main outcome measures. In broad terms, therefore, about two-thirds of Study 1 patients could be considered as substantially improved whereas only about one third of Study 2 patients could be categorized in this way. In addition, a significantly higher proportion of Study 2 patients reported receiving a lot of additional treatment over the follow-up period (3% in Study 1 *versus* 23% in Study 2).

Interpretation of this difference between studies is problematic since there were also differences in diagnostic selection criteria (1 month *v.* 6 months duration criterion), medication use (free of medication at entry to Study 1), clinical setting (primary *v.* secondary care), severity of baseline measures (greater in Study 2) demographic characteristics (significantly greater social deprivation in Study 2) and degree of interim treatment (greater in Study 2). One way of developing a perspective on the current results is to compare them with the follow-up study reported by Seivewright and her colleagues (1998) in which 20% of patients were described as having a poor or very poor outcome in terms of a standardized

assessment of health status based on case note review. They described their cohort of patients – a mixture of GAD, panic disorder and dysthymia – as intermediate between those seen in primary care only (as in Study 1) and those referred to outpatient psychiatric services (as in Study 2). Their sample shared some features with Study 1 (medication free at time of treatment, primary care setting, DSM-III diagnosis) and some features with Study 2 (referral to a psychiatric service, predominance of patients from social classes 4 and 5). Taken together the three cohorts of patients are probably best viewed as representing increasing degrees of complexity and severity in the broad spectrum of neurotic disorder seen in various clinical settings within the mental health services. Various factors influence the complexity and severity of clinical presentation in GAD and we are in the process of examining prognostic indicators in long-term follow-up studies of a number of clinical trials of CBT with anxiety disorders in central Scotland.

As well as differences in outcome between Studies 1 and 2 there were also similarities that are worth noting. The majority of patients in both studies remained symptomatic at long-term follow-up, irrespective of treatment received, and rates of recovery or remission were low (30–40%). Most patients nonetheless reported some degree of overall improvement probably because average levels of symptom severity were lower than pre-treatment levels. A sizeable minority of patients (30–40%), mainly from Study 2, had a generally poor outcome with chronic disability and high levels of continuing treatment in primary and secondary care. Taken together these results underline the therapeutic challenge of chronic conditions with complex clinical presentations (Durham *et al.* 2000). We are still some way from developing therapies for GAD that lead to full recovery and all the benefits that can flow from the remission of symptoms (cf. Ninian, 2001).

In light of the overall outcome in these two trials what evidence is there that CBT confers an advantage relative to other alternatives? In general, the CBT groups appeared to be better at long-term follow-up than the non-CBT groups. Only marginally significant differences were found when statistical tests were applied to patient classifications; the difference between the CBT and non-CBT groups on the BSI caseness measure approached 5% significance on a two-tailed test and a *post hoc* test showed that fewer patients from the CBT as opposed to the non-CBT group had received a lot of treatment in the follow-up period. The advantage of the CBT group was clearer when statistical tests were applied to quantitative measures combining data from the two studies. Looking at each study separately there were consistent differences in favour of CBT on all the measures of symptom severity and this pattern remained in analyses which took account of the missing data. Perhaps because of the small sample size the CBT advantage at long-term follow-up in Study 1 was not statistically significant. In Study 2 where there was less missing data a significant CBT advantage was found in both the observed and the missing data analyses. CBT, therefore, is associated with an attenuation of symptoms over the longer term but not with a reduced probability of remission.

Conclusion

The overall outcome was significantly better in Study 1 than Study 2. This was probably a function of the greater severity, chronicity and complexity of presenting problems in Study 2. The majority of patients in both studies remained symptomatic at long-term follow-up although average symptom severity was lower than pre-treatment levels and most reported some degree of improvement. For a sizeable minority (30–40%), mainly in Study 2, overall outcome was poor with evidence of continued disability and dependence on medication and/or psychological treatment. Although approximately 50% of patients in both CBT and non-CBT groups still had one or more clinical diagnoses at long-term follow-up, those in the CBT group had significantly lower levels of symptom severity on a number of psychometric measures. Finally, both CBT and the complexity and severity of presenting problems appear to influence the long-term outcome of GAD.

this project and to Donald Sharp, Mike Dow, Cathryn Hau and Andrew Gumley for their contribution as members of the Long-term Follow-up Research Team.

REFERENCES

Allison, P. D. (2001). *Missing data*. Sage: Thousand Oaks, California.

American Psychiatric Association (1980). *Diagnostic and Statistical Manual of Mental Disorders, 3rd edn*. APA: Washington DC.

American Psychiatric Association (1987). *Diagnostic and Statistical Manual of Mental Disorders, 3rd edn. Revised*. APA: Washington DC.

Ballenger, J. C., Davidson, J. R. T., Lecrubier, Y., Nutt, D. J., Borkovec, T. D., Rickels, K., Stein, D. J. & Wittchen, H.-U. (2001). Consensus statement on generalized anxiety disorder from the International Consensus Group on Depression and Anxiety. *Journal of Clinical Psychiatry* **62** (suppl. 11), 53–58.

Borkovec, T. D. & Ruscio, A. M. (2001). Psychotherapy for generalised anxiety disorder. *Journal of Clinical Psychiatry* **62** (suppl. 11), 37–42.

Brown, T. A., DiNardo, P. A. & Barlow, D. A. (1994). *Anxiety Disorders Interview Schedule for DSM-IV*. Graywind Publications: Albany, NY.

Carstairs, V. & Morris, R. (1991). *Deprivation and Health in Scotland*. Aberdeen University Press: Aberdeen.

Derogatis, L. R. (1993). *Brief Symptom Inventory (BSI). Administration, Scoring and Procedures Manual, 3rd edn*. National Computer Systems, Inc: Minneapolis.

Durham, R. C., Murphy, T. J., Allen, T., Richard, K., Treliving, L. & Fenton, G. W. (1994). A comparison of cognitive therapy, analytical psychotherapy and anxiety management training in the treatment of generalised anxiety disorder. *British Journal of Psychiatry* **165**, 315–323.

Durham, R. C., Swan, J. S. & Fisher, P. L. (2000). Complexity and collaboration in routine practice of CBT: what doesn't work with whom and how might it work better? *Journal of Mental Health* **9**, 429–444.

Fisher, P. L. & Durham, R. C. (1999). Recovery rates in generalized anxiety disorder following psychological therapy: an analysis of clinically significant change in STAI-T across outcome studies since 1990. *Psychological Medicine* **29**, 1425–1434.

Flick, S. N. (1988). Managing attrition in clinical research. *Clinical Psychology Review* **8**, 499–515.

Guy, W. (1982). Patient assessment in clinical trials. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **6** (suppl. 6), 601–606.

Hamilton, M. (1959). The assessment of anxiety states by rating. *British Journal of Medical Psychology* **32**, 50–55.

Hoehn-Saric, R., Borkovec, T. & Nemiah, J. (1995). Generalised anxiety disorder. In *Treatment of Psychiatric Disorders, 2nd edn.*

(ed. G. Gabbard), pp. 1537–1567. American Psychiatric Press: Washington, DC.

Jacobson, N. S. & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology* **59**, 12–19.

Jenkinson, C., Stewart-Brown, S., Petersen, S. & Paice, C. (1999). Assessment of the SF36 Version 2 in the UK. *Journal of Epidemiology and Community Health* **53**, 46–50.

Kellner, R. & Sheffield, B. F. (1973). A self-rating scale of distress. *Psychological Medicine* **3**, 88–100.

Kingdon, D., Tyrer, P., Seivewright, N., Ferguson, B. & Murphy, S. (1996). The Nottingham study of neurotic disorder: influence of cognitive therapists on outcome. *British Journal of Psychiatry* **169**, 93–97.

Luborsky, L., Diguer, L., Seligman, D., Rosenthal, R., Krause, E. D., Johnson, S., Halperin, G., Bishop, M., Berman, J. S. & Schweizer, E. (1999). The researcher's own therapy allegiances: a 'wild card' in comparisons of treatment efficacy. *Psychology Science and Practice* **6**, 95–106.

Ninian, P. T. (2001). Dissolving the burden of generalized anxiety disorder. *Journal of Clinical Psychiatry* **62** (suppl. 19), 5–10.

Ormel, J., Oldenhinkel, T., Brilman, E. & Brink, W. (1993). Outcome of depression and anxiety in primary care: a three wave $3\frac{1}{2}$ year study of psychopathology and disability. *Archives of General Psychiatry* **50**, 759–767.

Power, K. G., Jerrom, D. W. A., Simpson, R. J. & Mitchell, M. (1985). Controlled study of withdrawal symptoms and rebound anxiety after six weeks course of diazepam for generalised anxiety. *British Medical Journal* **290**, 1346–1248.

Power, K. G., Simpson, R. J., Swanson, V., Wallace, L. A., Feistner, A. T. C. & Sharp, D. (1990). A controlled comparison of cognitive-behaviour therapy, diazepam, and placebo, alone and in combination, for the treatment of generalised anxiety disorder. *Journal of Anxiety Disorders* **4**, 267–292.

Quinton, D., Gulliver, L. & Rutter, M. (1995). A 15–20 year follow-up of adult psychiatric patients. Psychiatric disorder and social functioning. *British Journal of Psychiatry* **167**, 315–323.

Roth, A. & Fonagy, P. (1996). *What Works for Whom? A Critical Review of Psychotherapy Research*. Guildford Press: London.

Seivewright, H., Tyrer, P. & Johnson, T. (1998). Prediction of outcome in neurotic disorder: a 5 year prospective study. *Psychological Medicine* **28**, 1149–1157.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R. & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y Self-Evaluation Questionnaire)*. Consulting Psychologists Press: Palo Alto, CA.

Tyrer, P. (2000). *Anxiety: A Multidisciplinary Review*. Imperial College Press: London.

Ware, J. E., Snow, K. K., Kosinski, M. & Gandek, B. (1993). *SF-36 Health Survey Manual and Interpretation Guide*. The Health Institute, New England Medical Center: Boston, MA.