British Journal of Clinical Psychology (2006), 45, 33–48

The British Psychological Society
www.bpsjournals.co.uk

Do empirically supported treatments generalize to private practice? A benchmark study of a cognitive-behavioural group treatment programme for social phobia

Jonathan E. Gaston[1,2]*, Maree J. Abbott[2], Ronald M. Rapee[2] and Sally A. Neary[1]

[1]Sydney Anxiety Disorders Practice, Sydney, Australia
[2]Macquarie University, Sydney, Australia

Objectives. There is much debate as to whether the treatment effects achieved in well-controlled studies such as randomized controlled trials (RCTs) are generalizable to more 'naturalistic' clinical populations, such as that seen in private practice. The current study sought to examine this issue in relation to social phobia.

Design. A benchmarking strategy was used to compare the effectiveness of a cognitive-behaviour therapy group programme for social phobia that was developed and evaluated in a research unit, to that of a private practice population.

Methods. Fifty-eight participants from a university research unit and 54 participants from an independent private practice who met the principal diagnostic criteria for social phobia completed the 10-session group programme. Symptom severity was measured at pre-treatment, post-treatment, and 3 months after treatment.

Results. No significant treatment differences were found between the research unit and private practice groups. Both groups showed significant treatment effects that were maintained at 3-month follow-up.

Conclusion. These findings suggest that treatments developed for RCTs are potentially transportable to private practice settings.

Social phobia involves a fear of social interaction or social performance situations in which there is the potential for embarrassment or scrutiny by others (American Psychiatric Association, 1994). Studies suggest that social phobia is both common and chronic in adult populations (Andrews, Henderson, & Hall, 2001; Jenkins *et al.*, 1997; Narrow, Rae, Robins, & Regier, 2002), with lifetime diagnostic estimates reaching over 13% (Kessler *et al.*, 1994). Social phobia can also be a highly disabling disorder, with

*Correspondence should be addressed to Jonathan Gaston, Department of Psychology, Macquarie University, Sydney, NSW 2109, Australia (e-mail: jgaston@psy.mq.edu.au).

DOI:10.1348/014466505X35146

often severe levels of life interference and impairment that are costly to both the individual and society (Mendlowicz & Stein, 2000; Rapee, 1995; Schneier *et al.*, 1994). Efficacious and empirically validated psychological treatments have been available for social phobia since the 1980s (Fedoroff & Taylor, 2001; Feske & Chambless, 1995; Gould, Buckminster, Pollack, Otto, & Yap, 1997; Taylor, 1996), showing good maintenance of treatment gains over the longer term (Heimberg, Salzman, Holt, & Blendell, 1993; Hunt & Andrews, 1998; Scholing & Emmelkamp, 1996). Despite the existence of these interventions, a range of national surveys have shown that less than 50% of people with social phobia receive any type of treatment, with rates of specialist treatment lower again (Andrews, Issakidis, & Carter, 2001; Bebbington, Brugha *et al.*, 2000; Bebbington, Meltzer *et al.*, 2000; Magee, Eaton, Wittchen, McGonagle, & Kessler, 1996). Of those who do access treatment services, many do not receive effective interventions and have poor longer-term outcomes (Andrews, Issakidis *et al.*, 2001; Bebbington, Meltzer *et al.*, 2000). A recent study by Issakidis and Andrews (2002) found that only 27.5% of people with social phobia in a national survey had received any specialist treatment and of these only 39% received a treatment considered efficacious.

For many individuals with social phobia, private practice represents a significant delivery point of these specialist treatment services, yet few data exist as to the quality and effectiveness of treatments delivered for any disorder in this type of setting (e.g. Clement, 1994). A number of authors have tried to explain why private practitioners do not evaluate the effectiveness of their interventions more routinely. Clement (1996) suggested that the lack of modelling of the integration of research and practice during clinical training, the lack of incentives and reinforcement for evaluating one's outcomes, and the lack of published research showing a positive impact of using evidence-based practice on referrals, retention, and income are all important factors. Other practical barriers may include the time needed to evaluate outcome, client resistance to evaluation, lack of suitable evaluation instruments, and lack of the necessary skills to conduct effective evaluation (Morrison, 1984). Practitioners are also often slow to upgrade their skills and implement newer and more effective treatments, even in the face of well-researched outcomes. This may be due to problems with the speed and effectiveness of dissemination, as well as practitioner resistance to the uptake of new clinical methods and treatments (Persons, 1995; Persons, 1997; Wilson, 1997). Practitioner resistance may be driven by a number of factors, including beliefs that regular clinical populations have a greater level of severity and co-morbidity compared with those used in randomized controlled trials (RCTs), and that treatments developed in these trials may not easily crossover to a whole range of specific clinical contexts (Barlow, Levitt, & Bufka, 1999). Wade, Treat, and Stuart (1998) have echoed these concerns, pointing out that controlled treatment outcome research commonly takes place under conditions that maximize both internal validity and the specificity of conclusions about causal mechanisms by using highly controlled and selective clinical environments. Havik and VandenBos (1996) have suggested that clinical populations in community settings are likely to be more heterogeneous, more severe, and have more problems overall. However, there is some disagreement as to whether the populations used in RCTs are necessarily as selective as assumed. In a review of the literature, Jacobson and Chistensen (1996) suggested that many RCTs are not exclusive and use populations with high severity and multiple problems. This claim is supported in a recent study by Stirman, DeRubeis, Crits-Christoph, and Brody (2003), who found that the reasons patients may not match with an RCT are less often due to issues of severity and complexity and more often due to patients having diagnoses that have not yet been

studied in the RCT literature. When severity is an issue, Stirman *et al.* found that it is more often the case that patients actually fail to meet the minimum severity criteria of an RCT. Hoagwood, Hibbs, Brent, and Jensen (1995) have suggested that other factors such as differential attrition rates and 'treatment drift' may also predict treatment variability across populations. Often treatment protocols in RCTs are manualized and strictly monitored, with an emphasis on maximizing treatment integrity. Treatments delivered in more naturalistic settings may not be as rigorous in terms of the content, quality or length of treatment. Related to this issue may be differences in therapist training, competencies, monitoring, and access to supervision (Merrill, Tolbert, & Wade, 2003). In an attempt to address these issues, authors such as Wilson (1996, 1997, 1998) have strongly argued for greater use of more standardized and manualized approaches in clinical settings.

Given the range of factors outlined above, to what extent can we assume that the treatment outcomes achieved in RCTs will generalize to other more 'naturalistic' settings such as private practice? One solution for establishing whether at least equal treatment outcomes can be achieved in a private practice setting as compared with controlled studies is the use of benchmarking (McFall, 1996). Benchmarking involves the establishment of reference points to interpret outcome data, whereby the magnitude of change obtained in efficacy studies can be used as a 'benchmark' against which change achieved in other clinical service settings can be judged.

To date, there has been a limited (although increasing) number of benchmarking studies specific to anxiety disorders. Empirically supported cognitive behavioural treatment (CBT) programmes for panic have been found to be efficacious across a number of different settings, including a regional outpatient clinic (Martinsen, Olsen, Tønset, Nyland, & Aarre, 1998), a community mental health centre (Wade *et al.*, 1998) and a public mental health unit (García-Palacios *et al.*, 2002). In terms of treating agoraphobia, Hahlweg, Fiegenbaum, Frank, Schroeder, and Witzleben (2001) found that the treatment effect sizes achieved in an outpatient clinic were comparable to that of RCTs. Comparable results have also been achieved for obsessive-compulsive disorder (OCD) in the same clinic setting using exposure therapy with response prevention (Wetzel, Bents, & Florin, 1999).

Few published studies have examined treatment generalization for social phobia to date. Haug *et al.* (2000) tested the effectiveness of exposure therapy for generalized social phobia when carried out by medical practitioners in a general practice setting. They found that exposure therapy alone, exposure plus medication, and medication alone were all significantly superior to placebo combined with general medical care. In a study using data gathered from four outpatient clinics in Germany, Lincoln *et al.* (2003) also found that individual exposure combined with cognitive restructuring for social phobia achieved comparable treatment effect sizes to that of controlled efficacy research. From a benchmarking perspective, however, there are a number of limitations to both of these studies. The Haug *et al.* (2000) study has been criticized as being too similar to an RCT, mainly due to their use of research-type recruitment strategies, randomization to treatment conditions, and overly selective inclusion criteria (Lincoln *et al.*, 2003). The Lincoln *et al.* study is also limited given that the treatment administered was not manualized or adopted from a pre-existing RCT and was not standardized in terms of overall treatment length or amount of exposure and cognitive therapy delivered to each client. Although this does not effect their overall conclusion regarding the success of using general cognitive behavioural strategies for social phobia in a community setting, it does restrict any conclusions that can be drawn about the

transportability of a treatment developed specifically for a RCT to that of a general clinical setting.

Even fewer studies have specifically examined benchmarking within private practice settings. Persons, Bostrom, and Beragnolli (1999) found that cognitive therapy for depression was as effective for patients treated in private practice as those achieved in RCTs. Persons *et al.* did find differences in their population compared with that of the RCTs. Generally, their sample was more heterogeneous in levels of depression (higher and lower levels of severity), had a greater number of major health problems, were more highly educated, and had less female members. In addition, more than twice the number of treatment sessions was used on average (34.8) in private practice compared with the RCTs. In a similar study of OCD treatment, Warren and Thomas (2001) found that a sample of private practice patients who completed cognitive behaviour therapy achieved comparable outcomes to that of representative RCTs. However, they found that their client sample was actually quite similar to that of RCT studies, with some small differences in gender ratio and duration of OCD prior to treatment.

Thus, there is emerging preliminary evidence that equivalent outcomes can be achieved in a private practice setting when compared with that of RCTs. Unfortunately, there have been no published benchmarking studies examining the effectiveness of social phobia treatment in private practice. Given the high prevalence and interference of this disorder, and the potential significance of private practice as a point of specialized treatment delivery, it would appear important to determine whether equivalent clinical outcomes can be achieved in private practice to that of RCTs. Therefore, the aim of the current study was to examine how well an empirically supported CBT group programme for social phobia transported to a private practice setting. Of particular interest was whether similar magnitudes of change could be achieved when compared with the university based clinical research unit in which the programme was developed and trialled.

Method

Participants

Two groups of individuals with social phobia participated in the study. The first group consisted of 58 participants who presented for assessment and treatment at the Macquarie University Anxiety Research Unit (MUARU), a clinical research unit that specializes in the assessment and treatment of anxiety disorders. Participants received subsidized assessment and free treatment in return for participating in research and an ongoing randomized treatment trial. The second group consisted of 54 individuals with social phobia who presented for assessment and treatment from the Sydney Anxiety Disorders Practice, an independent private clinical psychology practice that specializes in the assessment and treatment of anxiety disorders. Clients pay market rates for clinical services by experienced clinical psychologists. The research unit and private practice differed in their forms of participant recruitment as is typical for research and community populations. In general, MUARU recruited participants via media stories, self-referrals, or referrals from other professionals, whereas recruitment to the private practice was primarily through general practitioner referral or self-referral from pamphlet or website based advertising. There were few exclusion criteria for both settings; participants who met the principal diagnostic criteria for social phobia and who were not currently suicidal, self-harming, or experiencing psychosis were eligible for inclusion in the group programmes.

All participants met the *Diagnostic and Statistical Manual of Mental Disorders* criteria for a principal diagnosis of social phobia (DSM-IV; American Psychiatric Association, 1994). However, the method of determination of diagnoses differed between settings. Participants in the research unit programme were assessed by clinical psychology graduate students using the anxiety disorders interview schedule for DSM-IV (ADIS-IV; DiNardo, Brown, & Barlow, 1994). Clinical psychologists experienced in the assessment and treatment of adult anxiety disorders trained the graduate students in the structured interview. Inter-rater reliability for a principal diagnosis of social phobia using the ADIS-IV was calculated for the research unit using kappa coefficients, and showed excellent agreement ($\kappa = .89$). Avoidant personality disorder was diagnosed using the international personality disorder examination (ICD-10; Loranger, Janca, & Sartorius, 1997). Inter-rater reliability was also calculated for avoidant personality disorder diagnoses for this unit using kappa coefficients, and showed substantial agreement ($\kappa = .65$; Abbott, Peters, & Rapee, 2004). Clients of the private practice were diagnosed using a semi-structured clinical interview administered by clinical psychologists who were experienced in the assessment and treatment of anxiety disorders.

Measures

The questionnaire battery was completed at pre-treatment, post-treatment, and 3-month follow-up. Measures of social anxiety symptomatology included the Social interaction anxiety scale (SIAS; Mattick & Clarke, 1998), the Social phobia scale (SPS; Mattick & Clarke, 1998), and the Fear of negative evaluation scale, brief version (BFNE; Leary, 1983).

The SIAS was designed to measure fears related to social interaction defined as 'distress when meeting or talking with other people' (Mattick & Clarke, 1998, p. 457), whereas the SPS provides a measure of more specific fears of being scrutinized during regular activities (such as eating or drinking in public). Participants rate items on a 5-point scale from 'not at all true or characteristic of me' to 'extremely true or characteristic of me'. Total scores for both the SIAS and SPS range from 0 to 80, where a higher score indicates greater severity. Both scales have been shown to exhibit good reliability and validity and are sensitive to treatment change (Brown *et al.*, 1997; Heimberg, Mueller, Holt, Hope, & Liebowitz, 1992; Mattick & Clarke, 1998; Mattick & Peters, 1988; Mattick, Peters, & Clarke, 1989; Peters, 2000).

The BFNE measures fears of being evaluated negatively. Participants rate statements like, 'I am afraid that others will not approve of me' on a 5-point scale from 'not at all characteristic' to 'extremely characteristic'. Total BFNE scores range from 12 to 60, where a higher score indicates greater severity. While the BFNE exhibits good general psychometric properties (Leary, 1983; Rodebaugh *et al.*, 2004), it should be noted that there are currently no published studies examining its use with anxiety disorder populations.

The Physical and mental health summary scales (SF12; Ware, Kosinski, & Keller, 1995) were included as a measure of mental and physical health status, where lower scores indicate poorer health status. SF12 scores are standardized so that a mean of 50 represents the average score. The SF12 has been shown to be a reliable and valid measure of health status (e.g. Amir, Lewin-Epstein, Becker, & Buskila, 2002; Johnson & Coons, 1998). The SF12 Mental component scale has also been shown to be sensitive with anxious populations such that clinically anxious adults report high levels of mental health disability (Sanderson & Andrews, 2002; Sanderson, Andrews, & Jelsma, 2001).

The Depression anxiety stress scales (DASS; Lovibond & Lovibond, 1995) were designed to provide relatively distinct measures of depression, anxiety and stress. The DASS has been shown to exhibit good psychometric properties in both clinical and non-clinical populations (Antony, Bieling, Cox, Enns, & Swinson, 1998; Brown, Chorpita, Korotitsch, & Barlow, 1997; Clara, Cox, & Enns, 2001; Crawford & Henry, 2003; Lovibond, 1998; Lovibond & Lovibond, 1995). Participants in the research unit programme completed the 21-item short form of the DASS with trait wording, whereas those in the private practice completed the 42-item long-form with state wording. For both short and long versions, total scores for each subscale range from 0 to 42, where a higher score indicates greater severity.

Procedure

Treatment in both settings was equivalent. Participants attended a 10-session cognitive-behavioural group treatment programme for social phobia over a 12-week period, where the final sessions were staggered to allow time to practice and implement the skills taught. The programme was based on the approach of Rapee and Sanderson (1998), and utilized a therapist treatment manual that had previously been developed for an ongoing RCT at the research unit. The core components of this programme included realistic thinking, attention training, graded exposure, performance training and feedback, assertiveness training, and strategies to reduce perfectionism. Each group session lasted for approximately 2.5 hours and there was a maximum of eight participants in any group. Five therapists, including both registered clinical psychologists and clinical psychology graduate students, administered therapy in the research unit. Two registered clinical psychologists conducted treatment in the private practice, one of whom was also a therapist at the research unit. Participants completed the questionnaire battery before starting the programme, after completing the programme (i.e. at post-treatment) and again 3 months after completing the programme (i.e. 3-month follow-up). An additional group session was held at the 3-month follow-up point in order to assess progress for clients, although attendance at this session was voluntary.

Results

Demographic and diagnostic characteristics of the two samples

The two groups were compared on a number of demographic variables to assess for any significant differences at pre-treatment. Table 1 summarizes the age, gender, marital status, employment status, educational status, and medication characteristics for the two groups. The was no significant difference in mean age of participants in the research unit and private practice groups, $t(110) = 1.9, p > .05$. Chi-squared tests also showed no significant differences between the groups on gender distribution, $\chi^2(1) = 2.4, p > .05$, or marital status, $\chi^2(4) = 4.6, p > .05$. However, the two groups differed significantly in their education levels, $\chi^2(4) = 15.4, p < .01$, and employment $\chi^2(3) = 17.4, p < .01$[1]. The private practice sample had a greater proportion of participants who had completed an undergraduate degree and who were employed in managerial/professional positions than did the research unit group. Participant's medication use was coded according to whether it had been prescribed as anxiolytic or

[1] *Education and employment status were not significantly correlated with treatment change for the SIAS, SPS or the BFNE, all* rs $< .2$, p $> .05$.

antidepressant medication. The proportions of medication use are presented in Table 1 for the two groups, showing that almost one fifth of each group was prescribed antidepressant medication. Chi-squared analyses showed no significant differences in the proportion of participants in the two groups taking anxiolytic medication, $\chi^2(1) = 0.9, p > .05$, or antidepressant medication, $\chi^2(1) = 0.1, p > .05$.

Table 1. Demographic characteristics of the research unit and private practice groups

	Research unit (Mean or %)	Private practice (Mean or %)
Sample size	58	54
Age	34.8 (12.2)	31.1 (8.3)
Gender (% female)	51.7	37
% APD diagnosis	62.1	33.3
Marital Status		
Not married	70.7	61.1
Married	15.5	27.8
De facto	5.2	7.4
Separated	5.2	3.7
Divorced	3.4	0
Educational Status		
Less than high school	8.8	0
Completed high school	17.5	18.9
TAFE certificate	24.6	7.5
Undergraduate degree	42.1	71.7
Postgraduate degree	7	1.9
Employment Status		
Student/home-maker/unemployed	44.8	20.4
Trade/blue-collar work	15.5	9.3
Secretarial/administrative	15.5	7.4
Managerial/professional	24.1	63
Medication		
Anxiolytic	10.5	5.6
Antidepressant	24.6	22.2

All participants met the principal diagnostic criteria for social phobia (American Psychiatric Association, 1994). Of the research unit group, 96.4% were diagnosed with the generalized subtype of social phobia, and 94.4% of the private practice group were diagnosed likewise. There was no significant difference between the groups in the proportion of participants diagnosed with this subtype, $\chi^2(1) = 0.2, p > .05$. A greater proportion of the research unit group were given a diagnosis of avoidant personality disorder compared with the private practice group, 62.1% compared with 33.3%. Additional Axis 1 diagnoses for participants in the research unit group included generalized anxiety disorder (25.9%), other anxiety disorders (32.8%), dysthymia (25.9%), major depressive disorder (13.8%), and alcohol abuse/dependence (5.2%). Additional Axis 1 co-morbidity rates for the private practice sample included generalized anxiety disorder (27.8%), other anxiety disorders (3.7%), dysthymia (5.6%), major depressive disorder (11.1%), and alcohol abuse/dependence (3.7%). The mean number of additional co-morbid diagnoses for the research unit and private practice

groups was 1.1 (0.9) and 0.5 (0.7). Statistical analyses comparing Axis I and II comorbidity rates between the two groups were not conducted because a structured diagnostic interview (ADIS-IV) was not used for the private practice sample.

Comparison of pre-treatment symptom measures

The pre-treatment symptom measure scores for the two groups are presented in Table 2. Comparisons of these measures were made for the two groups using *t* tests and the experiment-wise error rate was controlled at $\alpha = .05$ using a Bonferroni correction to avoid inflation of the Type I error rate. No significant differences were found between the groups on pre-treatment SIAS scores, $t(110) = 1.8$, $p > .05$, SPS scores, $t(110) = 1.4$, $p > .05$, BFNE scores, $t(110) = 2.4$, $p > .05$, or on SF-12 physical component scale scores, $t(107) = 1.6$, $p > .05$, and mental component scale scores, $t(107) = 0.1$, $p > .05$.

Table 2. Completer mean symptom scores and (standard deviations) for the research unit and private practice groups at pre-treatment, post-treatment and follow-up

	Research unit			Private practice		
	Pre	Post	3mfu	Pre	Post	3mfu
SIAS	54.7 (11.9)	39.3 (13.4)	36.7 (13.9)	50.0 (15.5)	35.1 (17.0)	31.4 (16.4)
SPS	38.7 (14.5)	23.4 (12.8)	19.8 (10.0)	34.9 (14.5)	19.5 (15.4)	19.4 (16.6)
BFNE	51.3 (7.1)	42.9 (8.4)	40.3 (9.1)	47.7 (8.8)	38.6 (9.2)	37.8 (8.2)
SF12 – mental	34.3 (10.0)	42.9 (9.7)	42.8 (9.1)	34.1 (11.9)	45.2 (11.4)	45.1 (10.2)
SF12 – physical	51.8 (9.8)	54.1 (6.3)	53.3 (8.2)	54.5 (7.3)	55.5 (5.5)	55.1 (5.6)

The DASS depression scores for the two samples are not directly comparable because the instructions included differing wording (i.e. state vs. trait wording). However, it is of interest to note the clinical range of mean depression scores for each group separately. The pre-treatment trait DASS depression score for the research unit group fell in the moderate range (Mean $= 20.1$, $SD = 10.9$), whereas the state DASS depression score for the private practice group was in the mild range (Mean $= 13.0$, $SD = 11.4$).

Attrition rates

The mean number of sessions attended by the research unit and private practice groups was 7.8 (2.9) and 8.5 (1.5) sessions, respectively, showing no significant difference between the groups, $t(110) = 1.5$, $p > .05$. Participants who failed to attend 3 or more of the 10 sessions were considered to have discontinued treatment in that they had not completed enough of the programme for maximum benefit. Of the research unit group, 17% failed to attend eight or more sessions. In contrast, 7.1% of the private practice group failed to attend eight or more sessions. A chi-squared analysis showed that the attrition rate for the research unit group was significantly greater than that of the private practice, $\chi^2(1) = 4.9$, $p < .05$.

Questionnaire return rates

Overall, 75.9% of the research unit group and 85.2% of the private practice group returned postal and/or follow-up questionnaires after treatment. However, due to the relatively short period between post-treatment and follow-up periods (i.e. 3 months),

some participants only returned questionnaires for one of these data points. The number of participants in the research unit and private practice groups who completed postal questionnaires was 67.2% and 70.4%, respectively, while 63.8% of the research unit group and 68.5% of the private practice group completed 3-month follow-up questionnaires. Participants who returned questionnaires following treatment were compared with non-returners on measures of pre-treatment symptom severity and demographic variables. There were no significant differences with the exception of educational status, where a greater proportion of more highly educated participants returned questionnaires, $\chi^2(2) = 10.1, p < .05$.

Effects of treatment: Completer analyses

Treatment completers were defined as participants who had both attended at least eight group sessions and returned post-treatment questionnaires, giving 58.6% and 61.1% of the research unit and private practice groups, respectively. Repeated measures analyses of variance (ANOVA) were conducted to assess for any overall effects of treatment, any maintenance of treatment effects and any time × group interactions for each of the symptom measures. Two sets of planned contrasts were tested for each of the symptom measures. The first contrast compared symptom scores at pre- and post-treatment and the second compared scores at post-treatment and 3-month follow-up. Time × group interactions were also tested for each of these contrasts. The Completer means and standard deviations for each of the symptom measures at pre-, post-treatment and follow-up are reported in Table 2 for the two groups.

The present study hypothesized no difference in the size of treatment effects between the groups. It was considered inappropriate to correct for any inflation of the Type I error rate for contrasts assessing treatment effects, as this would bias the results in favour of the hypothesized absence of any group differences. The present analyses were therefore considered a more conservative test of the present hypotheses. Effect sizes (ES) were calculated using Cohen's d, where the ES = Mean1 − Mean2/(pooledstandarddeviation) for the completer and intent to treat analyses. Effect sizes were calculated for each of the symptom measures and are reported separately for the research unit (RU ES) and private practice (PP ES) groups.

Pre-versus post-treatment

There was a significant effect of time for the SIAS showing a reduction in scores at post-treatment, $F(1, 65) = 111.7$, $p < .001$, RU ES = 1.2; PP ES = 1.0. However, the time × group interaction for the SIAS was not significant, $F(1, 65) = 0.0$, $p > .05$. There was also a significant reduction in SPS scores from pre- to post-treatment, $F(1, 65) = 119.6$, $p < .001$, RU ES = 1.3; PP ES = 1.1, although the time × group interaction was not significant, $F(1, 65) = 0.0, p > .05$. Similarly, there was a significant effect of time for BFNE scores, $F(1, 65) = 89.1, p < .001$, RU ES = 1.2; PP ES = 1.2, but this effect did not differ across groups, $F(1, 65) = 0.5$, $p > .05$. There was a significant overall effect of time for the SF-12 mental component scale score, showing a reduction from pre- to post-treatment, $F(1, 62) = 39.2$, $p < .001$, RU ES = 0.8; PP ES = 1.0, but the time × group interaction was not significant, $F(1, 62) = 1.7$, $p > .05$. The SF-12 physical component score analysis did not show an effect of time, $F(1, 62) = 1.1, p > .05$, RU ES = 0.2; PP ES = 0.1, or any time × group interaction, $F(1, 62) = 0.6, p > .05$.

Pre- and post-DASS depression scores were compared using paired-samples t tests for each of the groups separately. The mean pre- and post-treatment DASS depression scores

were 19.7 (11.7) and 12.8 (6.8) for the research unit group and 12.2 (11.2) and 7.0 (7.8) for the private practice group, respectively. The results showed a significant reduction in DASS depression scores from pre- to post-treatment for the research unit group, $t(38) = 4.6$, $p < .001$, RU ES $= 0.7$, and the private practice group, $t(31) = 3.0$, $p < .01$, PP ES $= 0.6$.

Maintenance of treatment gains

In order to assess the maintenance of treatment gains for the two groups, a set of planned contrasts was tested comparing any effect of time from post-treatment to 3-month follow-up and any time × group interactions for each of the symptom measures.

Repeated Measures ANOVA showed a significant effect of time for the SIAS, such that scores at 3-month follow-up were significantly lower than at post, $F(1, 53) = 5.7$, $p < .05$, RU ES $= 0.2$; PP ES $= 0.2$, but the time × group interaction was not significant, $F(1, 53) = 0.4$, $p > .05$. Similarly, the BFNE showed a significant effect of time indicating a reduction in scores at 3-month follow-up, $F(1, 53) = 4.1$, $p < .05$, RU ES $= 0.3$; PP ES $= 0.1$, but the time × group interaction was not significant, $F(1, 53) = 0.9$, $p > .05$. The SPS showed an overall maintenance of gains from post-treatment to follow-up, $F(1, 53) = 2.6$, $p > .05$, RU ES $= 0.2$; PP ES $= 0.1$, and the size of this effect did not differ for the two groups, $F(1, 53) = 0.1$, $p > .05$. There was no significant effect of time, or time × group interactions for the SF-12 mental or physical component scale scores, all Fs < 1, *ns*, and effect sizes ranged from 0.0 to 0.1.

Effects of treatment: Intention-to-treat analyses

Intention to treat analyses were carried out to assess for any effects of treatment, any maintenance of treatment effects and any time × group interactions. Unlike the completer analyses, these analyses included all participants allocated to the study, regardless of whether they discontinued treatment or failed to return their postal or follow-up questionnaires. These analyses accounted for missing data by bringing forward the participant's last data score to replace the missing data point. The planned contrasts tested were the same as those in the completer analyses, that is, a comparison of symptom scores from pre- to post-treatment, comparisons of post-treatment and 3-month follow-up scores and any time × group interactions. The intention to treat mean scores for each of the symptom measures at pre-, post-treatment, and follow-up are reported in Table 3 for the two groups.

Table 3. Intention-to-treat mean symptom scores and (standard deviations) for the research unit and private practice groups at pre-treatment, post-treatment and follow-up

	Research unit			Private practice		
	Pre	Post	3mfu	Pre	Post	3mfu
SIAS	54.7 (11.9)	45.0 (14.5)	42.6 (15.7)	50.0 (15.5)	39.0 (17.4)	35.2 (17.0)
SPS	38.7 (14.5)	28.4 (14.8)	26.3 (15.4)	34.9 (14.5)	24.0 (15.2)	20.8 (15.3)
BFNE	51.3 (7.1)	45.8 (8.4)	43.5 (9.2)	47.7 (8.8)	40.5 (9.5)	39.0 (8.7)
SF12 – mental	34.3 (10.0)	39.5 (10.4)	40.4 (11.1)	34.1 (11.9)	42.11 (12.0)	42.3 (11.2)
SF12 – physical	51.8 (9.8)	52.9 (7.6)	52.6 (8.2)	54.5 (7.3)	54.5 (6.1)	54.5 (6.2)

Pre- versus post-treatment

The repeated measures ANOVA showed a significant reduction in SIAS scores from pre- to post-treatment, $F(1, 110) = 74.6, p < .001$, RU ES = 0.7; PP ES = 0.7, although the magnitude of this effect did not differ significantly for the two groups, $F(1, 110) = 0.3$, $p > .05$. Similarly, there was a significant reduction in SPS scores at post, $F(1, 110) = 78.7$, $p < .001$, RU ES = 0.7; PP ES = 0.8, but the time × group interaction was not significant, $F(1, 110) = 0.8$, $p > .05$. There was also a significant reduction in BFNE scores following treatment, $F(1, 110) = 64.2$, $p < .001$, RU ES = 0.7; PP ES = 0.8, but, again, no significant time × group interaction, $F(1, 110) = 1.0$, $p > .05$. The SF-12 mental component scale score was significantly lower at post-treatment, $F(1, 110) = 38.4$, $p < .001$, RU ES = 0.5; PP ES = 0.7, but this effect did not differ significantly across groups, $F(1, 110) = 1.8, p > .05$. The SF-12 physical component scale score did not show a significant effect of time, nor any time × group interaction, all Fs < 1, *ns*, RU ES = 0.0; PP ES = 0.1.

Maintenance of treatment gains

Each of the social phobia symptom measures showed significant improvements in overall mean scores from post-treatment to 3-month follow-up, SIAS: $F(1, 110) = 15.3$, $p < .001$, RU ES = 0.2; PP ES = 0.2; SPS: $F(1, 110) = 10.7, p < .01$, RU ES = 0.1; PP ES = 0.2; BFNE: $F(1, 110) = 14.7$, $p < .001$, RU ES = 0.3; PP ES = 0.2. However, none of these measures showed any significant time × group interactions, all Fs < 1, *ns*. There were no significant changes in mean SF-12 physical or mental component scale scores from post-treatment to follow-up, or any significant time × group interactions, all Fs < 1, *ns*, and effect sizes ranged from 0.0 to 0.1.

Discussion

The results of this benchmarking study of social phobia treatment suggest that comparable outcomes can be achieved in a private practice setting to that of an RCT in a university based clinical research unit. Not only were similar significant improvements found on all direct questionnaire measures of social phobia (SIAS, SPS, BFNE) from pre- to post-treatment, but both groups continued to improve at 3-month follow-up. The size of treatment effects from pre- to post-treatment ranged from 1.2 to 1.3 for the research unit, and 1.0–1.2 for the private practice. The magnitude of these effects compares favourably with that reported in meta-analytic studies of social phobia treatment, where the average effect sizes for cognitive-behavioural treatment range from .38 to 1.1 (Fedoroff & Taylor, 2001; Feske & Chambless, 1995; Gould *et al.*, 1997; Taylor, 1996). Similar patterns of improvement were also obtained at post-treatment for depression scores (DASS) and at post-treatment and 3-month follow-up on general mental health (SF-12), with no significant differences between the private practice and research unit populations. Encouragingly, the pattern of results across all measures was maintained even when using more conservative intention-to-treat analyses where treatment drop-out and missing data were taken into account.

A strength of this study is that it allows a direct comparison of private practice to a specific RCT using the same standardized treatment, instead of simply comparing the private practice outcome data to average effect sizes reported in meta-analytic studies. It is therefore possible to test the effect of sample characteristics on treatment generalizability more directly. On average, the private practice sample was more educated and had a higher employment status. They also appeared to have a lower

co-morbidity of other Axis I disorders and APD, although this difference could be due to different diagnostic procedures between the samples and was therefore not tested statistically. The private practice population also scored lower on depression, although again this difference could be due to using a state rather than trait version of the DASS. However, it should be stated that, overall, there were fewer differences between the research unit and private practice populations than might be anticipated. The finding that, if anything, the RCT sample may be more severe in terms of co-morbidity, levels of APD, and depression lends some support to the suggestion that many RCT populations are not exclusive and may have equal or even higher levels of severity and co-morbidity compared with more 'naturalistic' populations (Jacobson & Christensen, 1996; Stirman *et al.*, 2003). It is also encouraging to note that both the RCT and private practice populations had similar questionnaire return rates. This suggests that the ability and willingness of clients in a private practice setting to participate in clinical evaluation may not be as large an obstacle as has often been suggested (e.g. Morrison, 1984). It also supports the assertion by Persons (2001) that evidence-based practice is achievable in a private practice setting.

Interestingly, while there was no difference in treatment attendance, the private practice population actually had lower attrition (drop-out) rates than the RCT sample. The data therefore did not support Hoagwood *et al.*'s (1995) suggestion that differential attrition rates may lead to treatment variability across populations. Although the current study did not collect data pertaining to causes for differential attrition, one possible explanation is that the significant financial investment involved in private practice treatment may have lead to higher motivation and lower drop-out rates.

The current finding that the results achieved using an evidence-based treatment developed in a RCT can generalize to a more naturalistic setting such as private practice is consistent with studies looking at depression (Persons *et al.*, 1999) and OCD (Warren & Thomas, 2001). There are, however, a number of limitations of the current study. Firstly, the rates of completion from pre-to post-treatment for both groups are modest in size. Unfortunately, modest completion rates are common in the social phobia treatment literature (cf. Fedoroff & Taylor, 2001; Taylor, 1996). Furthermore, the higher attrition rate for the RCT may have potentially inflated the size of the treatment effect for this population. However, such an effect would be conservative if one expects the private practice programme to be less efficacious than the university programme. Therefore, the fact that the private practice population achieved similar treatment effect sizes increases the significance of achieving the RCT benchmark. In addition, the intention to treat analyses using last point carried forward is a very conservative means of correcting for differential attrition and these analyses still failed to show significant differences between the groups. It could also be argued that the populations are in fact too similar to be a true test of RCT generalizability. However, there is no evidence to suggest that the private practice population is different to that which 'naturalistically' occurs, as there were no specific recruitment or selection strategies used for this population other than social phobia being the primary problem. It could be that private practice clients are simply not as different to those found in RCT's as is often thought, or are, in fact, even more 'ideal' as a treatment population.

The current study does not answer questions as to the generalizability of RCT's to community health or hospital settings. It also does not address issues concerning the contribution of differences in therapist training and competence (Merrill *et al.*, 2003), as all clinicians involved in the current study had post-graduate level training in clinical psychology. The study is also limited by the fact that adherence to treatment protocols

was not measured in any standardized fashion to test for differences between therapist treatment delivery. However, the present study could be seen as responsive to Wilson's (1996, 1997, 1998) call for the greater use of more standardized and manualized approaches as a way of reducing potential therapist differences.

In conclusion, given the number of authors currently questioning the role and applicability of standardized and manualized evidence-based treatments to everyday clinical practice (e.g. Garfield, 1996; Havik & VandenBos, 1996), results of the current study provide evidence that these treatments can be effective in treating social phobia in a private practice setting. These results are also important as they suggest that it is possible to deliver efficacious and empirically validated psychological treatments to the significant number of people seeking help for social phobia in private practice settings.

References

Abbott, M. J., Peters, L., & Rapee, R. M. (2004). *Inter-rater reliability for social phobia and avoidant personality disorder in a large clinic sample.* (Unpublished manuscript).

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Amir, M., Lewin-Epstein, N., Becker, G., & Buskila, D. (2002). Psychometric properties of the SF-12 (Hebrew Version) in a primary care population in Israel. *Medical Care*, *40*, 918–928.

Andrews, G., Henderson, S., & Hall, W. (2001). Prevalence, comorbidity, disability and service utilisation: An overview of the Australian national mental health survey. *British Journal of Psychiatry*, *178*, 145–153.

Andrews, G., Issakidis, C., & Carter, G. (2001). The shortfall in mental health service utilisation. *British Journal of Psychiatry*, *179*, 417–425.

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment*, *10*, 176–181.

Barlow, D. H., Levitt, J. T., & Bufka, L. F. (1999). The dissemination of empirically supported treatments: A view to the future. *Behaviour Research and Therapy*, *37*, S147–S162.

Bebbington, P., Brugha, T., Meltzer, H., Jenkins, R., Ceresa, C., Farrell, M., & Lewis, G. (2000). Neurotic disorders and the receipt of psychiatric treatment. *Psychological Medicine*, *30*, 1369–1376.

Bebbington, P., Meltzer, H., Brugha, T., Farrell, M., Jenkins, R., Ceresa, C., & Lewis, G. (2000). Unequal access and unmet need: Neurotic disorders and the use of primary care services. *Psychological Medicine*, *30*, 1359–1368.

Brown, E. J., Turovsky, J., Heimberg, R. G., Juster, H. R., Brown, T. A., & Barlow, D. H. (1997). Validation of the Social Interaction Scale and the Social Phobia Scale across the anxiety disorders. *Psychological Assessment*, *9*, 21–27.

Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behaviour Research and Therapy*, *35*, 79–89.

Clara, I. P., Cox, B. J., & Enns, M. W. (2001). Confirmatory factor analysis of the Depression-Anxiety-Stress Scales in depressed and anxious patients. *Journal of Psychopathology and Behavioural Assessment*, *23*, 61–67.

Clement, P. W. (1994). Quantitative evaluation of 26 years of private practice. *Professional Psychology: Research and Practice*, *25*, 173–176.

Clement, P. W. (1996). Evaluation in private practice. *Clinical Psychology: Science and Practice*, *3*, 146–159.

Crawford, J. R., & Henry, J. D. (2003). The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology*, *42*, 111–131.

DiNardo, P. A., Brown, T. A., & Barlow, D. H. (1994). *Anxiety disorders interview schedule for DSM-IV.* San Antonio: Psychological Corporation, Harcourt Brace.

Fedoroff, I. C., & Taylor, S. (2001). Psychological and pharmacological treatments of social phobia: A meta-analysis. *Journal of Clinical Psychopharmacology*, *21*, 311–324.

Feske, U., & Chambless, D. L. (1995). Cognitive behavioral versus exposure only treatment for social phobia: A meta analysis. *Behavior Therapy*, *26*, 695–720.

García-Palacios, A., Botella, C., Robert, C., Baños, R., Perpiña, C., Quero, S., & Ballester, R. (2002). Clinical utility of cognitive-behavioural treatment for panic disorder. Results obtained in different settings: A research centre and a public mental health care unit. *Clinical Psychology and Psychotherapy*, *9*, 373–383.

Garfield, S. L. (1996). Some problems associated with validated forms of psychotherapy. *Clinical Psychology: Science and Practice*, *3*, 218–229.

Gould, R. A., Buckminster, S., Pollack, M. H., Otto, M. W., & Yap, L. (1997). Cognitive-behavioural and pharmacological treatment for social phobia: A meta-analysis. *Clinical Psychology: Science and Practice*, *4*, 291–306.

Hahlweg, K., Fiegenbaum, W., Frank, M., Schroeder, B., & Witzleben, I. (2001). Short- and long-term effectiveness of an empirically supported treatment for agoraphobia. *Journal of Consulting and Clinical Psychology*, *69*, 375–382.

Haug, T. T., Hellstrøm, K., Blomhoff, S., Humble, M., Madsbu, H. P., & Wold, J. E. (2000). The treatment of social phobia in general practice. Is exposure therapy feasible? *Family Practice*, *17*, 114–118.

Havik, O. E., & VandenBos, G. R. (1996). Limitations of manualized psychotherapy for everyday clinical practice. *Clinical Psychology: Science and Practice*, *3*, 264–267.

Heimberg, R. G., Mueller, G. P., Holt, C. S., Hope, D. A., & Liebowitz, M. R. (1992). Assessment of anxiety in social interaction and being observed by others: The Social Interaction Anxiety Scale and the Social Phobia Scale. *Behavior Therapy*, *23*, 53–73.

Heimberg, R. G., Salzman, D. G., Holt, C. S., & Blendell, K. (1993). Cognitive behavioral group treatment for social phobia: Effectiveness at five-year follow-up. *Cognitive Therapy and Research*, *17*, 325–339.

Hoagwood, K., Hibbs, E., Brent, D., & Jensen, P. (1995). Introduction to the special section: Efficacy and effectiveness in studies of child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology*, *63*, 683–687.

Hunt, C., & Andrews, G. (1998). Long-term outcome of panic disorder and social phobia. *Journal of Anxiety Disorders*, *12*, 395–406.

Issakidis, C., & Andrews, G. (2002). Service utilisation for anxiety in an Australian community sample. *Social Psychiatry and Psychiatric Epidemiology*, *37*, 153–163.

Jacobson, N. S., & Christensen, A. (1996). Studying the effectiveness of psychotherapy: How well can clinical trials do the job? *American Psychologist*, *51*, 1031–1039.

Jenkins, R., Lewis, G., Bebbington, P., Brugha, T., Farrell, M., Gill, B., & Meltzer, H. (1997). The National Psychiatric Morbidity Surveys of Great Britain – initial findings from the household survey. *Psychological Medicine*, *27*, 775–789.

Johnson, J. A., & Coons, S. J. (1998). Comparison of the EQ-5D and SF-12 in an adult US sample. *Quality of Life Research*, *17*, 155–166.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H. U., & Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Survey. *Archives of General Psychiatry*, *51*, 8–19.

Leary, M. (1983). A brief version of the fear of negative evaluation scale. *Personality and Social Psychology Bulletin*, *9*, 371–375.

Lincoln, T. M., Rief, W., Hahlweg, K., Frank, M., Witzlebel, I., Schroeder, B., & Fiegenbaum, W. (2003). Effectiveness of an empirically supported treatment for social phobia in the field. *Behaviour Research and Therapy*, *41*, 1251–1269.

Loranger, A. W., Janca, A., & Sartorius, N. (1997). *Assessment and diagnosis of personality disorders: The ICD-10 international personality disorder examination (IPDE)*. Cambridge, UK: Cambridge University Press.

Lovibond, P. F. (1998). Long-term stability of depression, anxiety, and stress syndromes. *Journal of Abnormal Psychology*, *107*, 520–526.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck depression and anxiety inventories. *Behaviour Research and Therapy*, *33*, 335–342.

Magee, W., Eaton, W., Wittchen, H. U., McGonagle, K., & Kessler, R. (1996). Agoraphobia, simple phobia and social phobia in the National Comorbidity Survey. *Archives of General Psychiatry*, *53*, 159–168.

Martinsen, E. W., Olsen, T., Tønset, E., Nyland, K. E., & Aarre, T. F. (1998). Cognitive-behavioural group therapy for panic disorder in the general clinical setting: A naturalistic study with 1-year follow-up. *Journal of Clinical Psychiatry*, *59*, 437–442.

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*, *36*, 455–470.

Mattick, R. P., & Peters, L. (1988). Treatment of severe social phobia: Effects of guided exposure with and without cognitive restructuring. *Journal of Consulting and Clinical Psychology*, *56*, 251–260.

Mattick, R. P., Peters, L., & Clarke, J. C. (1989). Exposure and cognitive restructuring for severe social phobia: A controlled study. *Behavior Therapy*, *20*, 3–23.

McFall, R. M. (1996, July). *Consumer satisfaction as a way of evaluating psychotherapy: Ecological validity and all that versus the good old randomized trial (panel discussion)*. Paper presented at the 6th Annual Convention of the American Association of Applied and Preventative Psychology, San Francisco.

Mendlowicz, M. V., & Stein, M. B. (2000). Quality of life in individuals with anxiety disorders. *American Journal of Psychiatry*, *157*, 669–682.

Merrill, K. A., Tolbert, V. E., & Wade, W. A. (2003). Effectiveness of cognitive therapy for depression in a community mental health centre: A benchmarking study. *Journal of Consulting and Clinical Psychology*, *71*, 404–409.

Morrison, J. K. (1984). Crossing the research barrier in the private practice of psychotherapy. *Psychotherapy in Private Practice*, *11*, 37–42.

Narrow, W. E., Rae, D. S., Robins, L. N., & Regier, D. A. (2002). Revised prevalence estimates of mental disorders in the Unites States. *Archives of General Psychiatry*, *59*, 115–123.

Persons, J. B. (1995). Why practicing psychologists are slow to adopt empirically-validated treatments. In S. C. Hayes, V. M. Follette, R. M. Dawes, & K. E. Grady (Eds.), *Scientific standards of psychological practice: Issues and recommendations* (pp. 12–44). Reno, Nevada: Context Press.

Persons, J. B. (1997). Dissemination of effective methods: Behavior therapy's next challenge. *Behavior Therapy*, *28*, 465–471.

Persons, J. B. (2001). Conducting effectiveness studies in the context of evidence-based clinical practice. *Clinical Psychology: Science and Practice*, *8*, 168–172.

Persons, J. B., Bostrom, A., & Bertagnolli, A. (1999). Results of randomized controlled trials of cognitive therapy for depression generalize to private practice. *Cognitive Therapy and Research*, *23*, 535–548.

Peters, L. (2000). Discriminant validity of the Social Phobia and Anxiety Inventory (SPAI), the Social Phobia Scale (SPS) and the Social Interaction Scale (SIAS). *Behaviour Research and Therapy*, *38*, 943–950.

Rapee, R. M. (1995). Descriptive psychopathology of social phobia. In R. G. Heimberg, M. R. Liebowitz, D. A. Hope & F. R. Schneier (Eds.), *Social phobia: Diagnosis, assessment and treatment* (pp. 41–66). New York: Guilford Press.

Rapee, R. M., & Sanderson, W. C. (1998). *Social Phobia: Clinical application of evidence-based psychotherapy.* New Jersey: Jason Aronson.

Rodebaugh, T. L., Woods, C. M., Thissen, D. M., Heimberg, R. G., Chambless, D. L., & Rapee, R. M. (2004). More information from fewer questions: The factor structure and item properties of the original and brief fear of negative evaluation scale. *Psychological Assessment*, *16*, 169–181.

Sanderson, K., & Andrews, G. (2002). Prevalence and severity of mental health-related disability and relationship to diagnosis. *Psychiatric Services*, *53*, 80–86.

Sanderson, K., Andrews, G., & Jelsma, W. (2001). Disability measurement in the anxiety disorders: Comparison of three brief measures. *Journal of Anxiety Disorders*, *15*, 333–344.

Schneier, F. R., Heckelman, L. R., Garfinkel, R., Campeas, R., Fallon, B. A., Gitow, A., Street, L., Del Bene, D., & Liebowitz, M. R. (1994). Functional impairment in social phobia. *Journal of Clinical Psychiatry*, *55*, 322–331.

Scholing, A., & Emmelkamp, P. M. G. (1996). Treatment of generalized social phobia: Results at long-term follow-up. *Behaviour Research and Therapy*, *34*, 447–452.

Stirman, S. W., DeRubeis, R. J., Crits-Christoph, P., & Brody, P. E. (2003). Are samples in randomized controlled trials of psychotherapy representative of community outpatients? A new methodology and initial findings. *Journal of Consulting and Clinical Psychology*, *71*, 963–972.

Taylor, S. (1996). Meta-analysis of cognitive-behavioral treatments for social phobia. *Journal of Behavior Therapy and Experimental Psychiatry*, *27*, 1–9.

Wade, W. A., Treat, T. A., & Stuart, G. L. (1998). Transporting an empirically supported treatment for panic disorder to a service clinic setting: A benchmarking strategy. *Journal of Consulting and Clinical Psychology*, *66*, 231–239.

Ware, J. E., Kosinski, M. A., & Keller, S. D. (1995). *SF-12: How to score the SF-12 Physical and Mental Health Summary Scales* (2nd ed). Boston, MA: The Health Institute, New England Medical Center.

Warren, R., & Thomas, J. C. (2001). Cognitive-behavior therapy of obsessive-compulsive disorder in private practice: An effectiveness study. *Journal of Anxiety Disorders*, (15), 277–285.

Wetzel, C., Bents, H., & Florin, I. (1999). High-density exposure therapy for obsessive-compulsive inpatients: A 1-year follow-up. *Psychotherapy and Psychosomatics*, *68*, 186–192.

Wilson, G. T. (1996). Empirically validated treatments: Reality and resistance. *Clinical Psychology: Science and Practice*, *3*, 241–244.

Wilson, G. T. (1997). Dissemination of cognitive behavioral treatments. *Behavior Therapy*, *28*, 473–475.

Wilson, G. T. (1998). Manual-based treatment and clinical practice. *Clinical Psychology: Science and Practice*, *5*, 363–375.

Received 19 January 2004; revised version received 19 August 2004