NORMATIVE VALUES FOR THE BECK ANXIETY INVENTORY, FEAR QUESTIONNAIRE, PENN STATE WORRY QUESTIONNAIRE AND SOCIAL PHOBIA AND ANXIETY INVENTORY

by

Martha M. Gillis

submitted to the

Faculty of the College of Arts and Sciences

of The American University

in Partial Fulfillment of

the Requirements for the Degree

of

Master of Arts

in

Psychology

Signatures of Committee:

Chair: D. A. F. Haynes

Clinical Assessment

Gary T. Ford

Betty T Bennett

Dean of the College

5-5-94

Date

1994

The American University

Washington, D.C. 20016

NORMATIVE VALUES FOR THE BECK ANXIETY INVENTORY, FEAR QUESTIONNAIRE, PENN STATE WORRY QUESTIONNAIRE AND SOCIAL PHOBIA AND ANXIETY INVENTORY

by

Martha M. Gillis

Abstract

Assessment of the meaningfulness or magnitude of therapeutic change (known as clinical significance), while believed to be highly desirable, even necessary, has been hampered by a lack of normative data on many therapy outcome measures. In this study, community data are reported for four psychometrically sound measures widely used in psychotherapy outcome research: the Beck Anxiety Inventory, the Fear Questionnaire, the Penn State Worry Questionnaire, and the Social Phobia and Anxiety Inventory. The demographic profile of the community samples closely matches the 1990 U.S. national census profile on age, race, income, and gender. Subgroup comparisons were made between Blacks and Caucasians, between the lowest income quintile and the other income groups, and between men and women. Means and standard deviations and percentile scores are provided for each measure with examples of their use in several methods for calculating clinical significance.

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

Evaluations of the efficacy of treatments for psychological disorders have traditionally asked whether a specific treatment led to improvements in client functioning or which of several comparison treatments led to greater improvement. With the exception of single-case studies, the answers have generally been sought through statistical analyses at the group level, using pre- to posttreatment repeated measures comparisons or between-group posttreatment comparisons.

Because such studies are most frequently conducted with small samples that have only weak power to detect group differences when comparing treatments (Kazdin & Bass, 1989), large improvements (or deterioration) by a small number of clients can be masked by non-significant statistical differences between groups. The clinical, ethical, and theoretical implications would be quite different for treatments occasionally producing significant deterioration versus those occasionally producing significant improvement or that were overall only marginally effective.

More importantly, this approach fails to address the magnitude or meaningfulness of improvement. Meaningfulness of improvement is linked to a treatment's ability to

effectively meet therapeutic goals, such as producing posttreatment functioning that is normal or near-normal. This type of improvement has come to be known as clinically significant change. The two concepts--statistical significance and clinical significance--are not equivalent (Hugdahl & Ost, 1981; Jacobson & Truax, 1991; Wampold & Jenson, 1986). Individuals and groups with a high degree of pathology prior to treatment can improve to a statistically significant level without approaching normal functioning (e.g., Beidel, Turner & Cooley, 1993). Yet the distinction is not always made in the literature (Jacobson & Truax, 1991).

CHAPTER 1

STATEMENT OF THE PROBLEM

Jacobson, Follette and Revenstorf (1984) offered three alternative criteria for establishing the clinical significance of individual subjects' change: 1) a posttreatment score should fall two standard deviations beyond the mean of the dysfunctional population (in the direction of improvement); 2) the posttest score should fall within two standard deviations of the mean of the functional ("normal") population; or 3) the posttest score should be more likely statistically to be from the functional rather than the dysfunctional population.

The first proposed criterion is an arbitrary statistical solution which, while requiring large improvement, still fails to establish where that improvement lies relative to normality; individuals in more dysfunctional groups might possibly achieve a two-standard-deviation improvement while continuing to exhibit considerable pathology (Jacobson & Truax, 1991; Wampold & Jenson, 1986). The second and third proposals overcome that problem, but the second may be too stringent (e.g. Turner et al., 1993) and both raise fundamental issues over how to define normal (e.g., Blanchard & Schwarz, 1988;

Hollon & Flick, 1988; Jacobson & Revenstorf, 1988; Kendall & Grove, 1988; Saunders, Howard & Newman, 1988).

Defining Normal

The three most common definitions of normal are average, healthy, or optimal functioning. Normal as average usually means typical. For example, one might say that the average American is 10 pounds overweight. On the other hand, healthy functioning is defined as the absence of symptoms (not overweight, not sick), while optimal functioning implies some level of excellence or well-being (not only not overweight, but flexible and physically fit with an ideal body/fat ratio).

Jacobson et al.'s (1984) third criterion--that a posttest score would be statistically more likely to be from the functional rather than the dysfunctional population--assumes that dysfunctional people represent a separate population from functional ones. This calls for either the healthy or the optimal definition. In situations where the goal of treatment is complete elimination of a behavior, such as smoking (Baer, 1988) or self-destructiveness (suicidality), healthy or optimal functioning might be the preferred definition. Since health risks associated with smoking are cumulative, for example, smoking any number of cigarettes, even a number lower than the average for the population as a whole, could be considered clinical failure. The desirable normative comparison in such cases would be a

completely symptom-free population (healthy functioning--non-smokers).

Defining normal as average, on the other hand, may be more suitable in relation to target behaviors one does not want/need to eliminate entirely, or whenever different contexts make different symptom levels adaptive. For example, preoccupation with somatic symptoms is associated with psychopathology in physically healthy individuals, but is adaptive in certain chronically ill people, such as diabetics. Likewise, a level of vigilance that reflects paranoia in the average case may be common sense in an inner city drug zone or Sarajevo. Such atypical subjects will appear more statistically extreme compared to an asymptomatic normative criterion than they would to an average one because averages are sensitive to extreme values and are pulled in their direction. The mean of a sample that includes culturally "normal," but statistically extreme values will therefore be closer to that extreme than the mean of a sample that excludes such values.

Eliminating this problem completely would require establishing subgroup norms, which in turn requires a priori definition of subgroups and their characteristics. The number of such groups is potentially infinite, and identification of their defining characteristics is problematic (Kazdin & Wilson, 1978). One can reduce the problem, however, by defining normal as average and sampling

the full range of behavior associated with dysfunctionality and functionality.

There are also methodological reasons to sample the full range of behavior. Truncating the distribution by eliminating dysfunctional scorers restricts the range of scores. In addition, attempting to delineate two populations by dichotomous categorization of subjects as functional/dysfunctional on the basis of a cutoff score is subject to measurement error, resulting in misclassifications. Finally, sampling the full range of behavior has the advantage that it does not require any a priori definition or identification of dysfunctionality.

Alternative Methods for Calculating Clinical Significance

Even if the functional and dysfunctional populations are separate, to the degree that they overlap, the two-standard-deviation cutoff in both the first and the second criteria proposed by Jacobson et al. (1984) may be excessively stringent. Hollon and Flick (1988) argue against using the two-standard-deviation cutoff, on the additional grounds that it is completely arbitrary and imposes a categorical judgment on what may be a continuously distributed variable. Instead, they suggest reporting percentages of patients falling one, two, and three or more standard deviations from the mean of a normative sample. It can be argued, however, that this suggestion is simply a more extensive, but equally arbitrary, categorization

scheme. Comparison of posttreatment scores of individual subjects to the mean score of normal subjects on the same measures, however, does not have to be reported in terms of standard deviations at all, a format that Jacobson et al. (1984) themselves acknowledge may be least appealing to clinicians. Jacobson and Truax (1991) note that, while the need for determining clinical significance has gained acceptance, which method best establishes that significance currently involves subjective decisions.

The Need for Norms

Whether one adopts the Jacobson et al. or the Hollon and Flick method, or something else entirely, such as percentiles or a composite measure suggested by Turner et al. (1993) in which clinical significance must be achieved on 80% of the outcome measures used, all the proposed methods have in common the requirement that scores of normal subjects be available and that an operational definition of normal be selected and specified.

CHAPTER 2

METHOD

The Design

Because anxiety disorders are among the most common disorders treated in psychotherapy (e.g. American Psychiatric Association, 1987; Boyd et al., 1990; Regier, Narrow & Rae, 1990; Reich, 1986), and there is a growing body of research into the relative effectiveness of a variety of treatments for them (e.g. Borkovec & Mathews, 1988; Chambless & Gillis, 1993; Heimberg, 1989), this study is designed to provide normative information for several common psychotherapy outcome measures utilized for generalized anxiety disorder, social phobias, agoraphobia, and other anxiety disorders. The measures selected were the Beck Anxiety Inventory ([BAI] Beck, Epstein, Brown & Steer, 1988), Fear Questionnaire ([FQ] Marks & Mathews, 1979), Penn State Worry Questionnaire ([PSWQ] Meyer, Miller, Metzger & Borkovec, 1990), and Social Phobia and Anxiety Inventory ([SPAI] Turner, Beidel, Dancu & Stanley, 1989). These instruments are all self-report measures selected on the basis of their empirically established reliability and validity, appropriateness for psychotherapy outcome research, and, with the exception of the Fear Questionnaire

(FQ), absence of established adult community norms.

Subjects

Two convenience samples (n = 228, n = 245) were selected on a quota basis to match the demographic profile of U.S. adults between the ages of 18 and 65 (U.S. Department of Commerce, 1990) across four variables: age, race, gender, and income. The mean age for subjects in both samples was 37 years, 52% were women, 12% Black, and 84% Caucasian. Hispanic Caucasians represent 9% of the total samples. See Tables 1 and 2 for further demographic information.

For the purposes of this study, normal was defined as average. Therefore no effort was made to screen for psychopathology.

Originally only mall intercepts were planned, but it proved extremely difficult to locate both Hispanics and lower income (below $28,000) subjects at those locations. As a result, potential subjects were approached at a Catholic church, a trailer park, and a homeless shelter, in addition to two shopping malls, all in the suburbs outside Washington, D.C. Most of the subjects (n = 435) were offered the incentive of receiving a free District of Columbia lottery ticket in return for agreeing to complete the questionnaires. Many of the Hispanic subjects recruited through the church did not read English (SPAI n = 10 of 21; BAI, PSWQ, and FQ, n = 19 of 25), so the questions were

orally translated into in Spanish for them, but they wrote their own responses. A $10 donation was made to the church for each questionnaire this group completed. The demographic profile for subjects responding to each of the questionnaires is displayed in Table 1 together with the target national census quotas. The numbers reported for each of the questionnaires differ slightly, partly because the data of subjects who left blank or answered incorrectly (e.g., two numbers circled) more than 10% of the questions were excluded from analyses for that questionnaire. This was an arbitrary cutoff for exclusion of data whose validity seemed questionable.

In addition, a few subjects gave an answer during the oral screening that differed from the written answer(s) they provided on the demographic questionnaire. In cases where the discrepancy between the oral and written answers was inexplicable, suggesting possible misrepresentation or interviewer error, the data for those subjects were excluded from analyses. The percentages in the final samples thus differ slightly from one another and from the target quotas (see Table 1). None of the differences are significant. Although more Blacks than Caucasians and more women than men were excluded from data analyses, chi-square analyses indicate there were no significant differences between the excluded subjects and those whose data were included on any of the four quota variables.

In some cases the family income answer supplied on the demographic form failed to match the family income range given during oral screening but did match the personal income range given during screening. Subjects with this type of discrepancy were not excluded from analyses, but the analyses were based on the family income reported during screening rather than that reported on the demographic form.

Some subjects omitted answers to one or more questions on the demographic form. This may be linked to comments by some subjects, reported by the marketing representatives, that questions asked during screening were repeated on the demographics form and suggesting that the subjects found providing duplicate answers unnecessary or burdensome.

Subjects with incomplete demographic forms were not excluded from analyses, provided that the critical quota information was available either from the form or from the screening information. Because the data from the demographic form are more specific and cover more variables, wherever data were available from both the screening and the demographic form, except as already noted, the descriptive statistics (Table 2) and paired subgroup analyses are based on data from the demographic form.

Procedure

All the potential subjects except those recruited through the church were approached by representatives of two professional marketing firms hired by the investigator.

They were asked if they were willing to answer some questions. Screening questions on the quota variables were asked first. Subjects who did not meet selection criteria were thanked and the interview terminated, while subjects who met selection criteria were asked if they were willing to take ten minutes to fill out some questionnaires.

Pilot data indicated that the Social Phobia and Anxiety Inventory (SPAI) takes approximately ten minutes to complete. Therefore, in order to avoid high refusal/incompletion rates and to avoid overburdening individual subjects, the SPAI alone was administered to one sample. The other sample received the Beck Anxiety Inventory, Penn State Worry Questionnaire, and the Fear Questionnaire. Prior to completing the questionnaires, all subjects received a letter explaining the purpose of the study and its voluntary nature, and providing telephone numbers to call if any concerns arose subsequent to completion of the packet. The last item of the packet for all subjects was a brief questionnaire that asked for more specific information on the screening variables and for some additional demographic data.

The BAI, PSWQ, and FQ were packaged in random order to control for possible response-order effects.

Anonymity was guaranteed by not having any identifying information on the questionnaires and by foregoing signed consent.

Measures

Fear Questionnaire (FQ)

The 23 items of the FQ (Marks & Mathews, 1979) include several subscales, among them one for social phobia and one for agoraphobia. Scores for these subscales are obtained when subjects indicate to what degree they avoid 15 specified situations or places because of fear, using a scale of 0-8 ("Would not avoid it" - "always avoid"). For five additional items comprising an anxiety/depression subscale, subjects indicate how troublesome specific symptoms are using a similar 0-8 scale ("Hardly at all" - "very severely troublesome"). For clinical use, there are three blank items where subjects can write in and rate the phobic situation for which they are seeking treatment or any avoided/troublesome situation/symptom not included in the standardized list.

Mavissakalian (1986) reports that correlations between the Agoraphobia subscale and the other subscales were .28 or less, and that pre- to posttreatment changes in the Agoraphobia subscale discriminated agoraphobic subjects' rates of improvement as validated by external criteria such as global clinician ratings and performance on a behavior test. The FQ Agoraphobia subscale has a 1-week test-retest reliability of .89, while the Social Phobia subscale has a 1-week test-retest reliability of .82 (Marks & Mathews, 1979).

The factor structure underlying the FQ subscales has been examined and supported (Lelliott, McNamee, & Marks, 1991; Oei, Moylan, & Evans, 1991). In addition, the ability of its agoraphobia and social phobia subscales to discriminate subjects diagnosed with those two disorders from each other, as well as from other anxiety patients has been supported in several studies (Cox, Swinson & Shaw, 1991; Lelliott, McNamee, & Marks, 1991; Oei, Moylan, & Evans, 1991).

The FQ's assessment of fear and avoidance combined, rather than separately, has been validated by Cox, Swinson and Fergus (1993). Although sex differences in response patterns to the FQ have been detected (e.g. Lelliott, McNamee, & Marks, 1991; see Moylan & Oei, 1992 for a review), Arrindell and Buikhuisen (1992) report that, with the exception of the Blood-Injury subscale (not used in this study), these results are not explained by social demand factors and that in general the FQ seems unaffected by social desirability response bias.

For the FQ, two studies to establish community norms have already been conducted (Mizes & Crawford, 1986; Nietzel & Trull, 1988; Trull & Hillerbrand, 1990). However, a marked discrepancy between the means obtained in the two studies needs to be resolved. Trull and Hillerbrand (1990) reported a community mean of 33.0 (SD = 13.1) for men and 46.1 (SD = 16.2) for women on Total Phobia (possible range =

0-160), 14.2 (SD = 4.3) for men and 16.3 (SD = 6.2) for women on the Social Phobia subscale (possible range = 0-48), and 7.9 (SD = 7.1) for men and 14.8 (SD = 8.5) for women on the Agoraphobia subscale (possible range = 1-48). Mizes and Crawford, on the other hand, report means of 23.9 (SD = 15.9) for female adults and 26.1 (SD = 16.2) for male adults on the Total Phobia scale, with a female mean of 5.0 (SD = 6.0) on the Agoraphobia subscale and a male mean of 4.9 (SD = 5.1). Their Social Phobia subscale mean was 8.8 (SD = 6.1) for female adults and 8.9 (SD = 5.5) for men. These results all exclude the three write-in items. Moreover, the two studies' findings with regard to subgroup differences were reversed. Trull and Hillerbrand (1990) found that community subjects reported significantly greater levels of fear than collegiate subjects did and that women were more fearful than men. Mizes and Crawford (1986), on the other hand, found lower levels of fear among their adult community sample than their collegiate sample, and only among the students were women more fearful than men.

Since the Trull and Hillerbrand (1990) study was conducted in an urban area (Lexington, Kentucky) by telephone, while the Mizes and Crawford (1986) study was conducted in a smaller, midwestern community (Fargo, North Dakota) using a mailed, written format, the discrepant findings may be the result of either methodological or regional differences. Trull and Hillerbrand (1990),

therefore, recommended collection of normative data from an additional large metropolitan area. To accomplish that goal, for this study the FQ was administered in the greater Washington, D.C., metropolitan area, using written format and quota sampling, predominantly through mall intercepts.

The question of generalizability between communities is particularly relevant for the anxiety disorders, since the National Institute of Mental Health (NIMH) Epidemiologic Catchment Area (ECA) study results suggest that there are regional differences in the prevalence of phobias (including agoraphobia) which have not been satisfactorily explained by variations in intersite demographics or methodology (Robins et al., 1984).

Beck Anxiety Inventory (BAI)

The BAI (Beck, Epstein, Brown & Steer, 1988) consists of 21 items describing anxiety symptoms. Respondents are asked to rate how much each symptom bothered them during the previous week, using a 4-point scale ranging from 0 (Not at all) to 3 (Severely--I could barely stand it).

The inventory was developed in response to evidence that existing anxiety measures had poor ability to discriminate between depression and anxiety, and it has better ability to make that discrimination than the State-Trait Anxiety Inventory (STAI), Zung Self-Rating Scale, Taylor Manifest Anxiety Scale, and Hamilton Anxiety Rating Scale-Revised (Beck et al., 1988). The BAI's developers

reported a one-week test-retest reliability of .75 and high internal consistency (alpha = .92). Fydrich, Dowdall and Chambless (1992) reported a test-retest reliability (avg. = 11 days) of .67, internal consistency of .94, "robust convergent validity" (p. 59), and better discriminant validity than the STAI. Beck and Steer (1991) also found support for the concurrent and discriminant validity of the BAI.

Penn State Worry Questionnaire (PSWQ)

The PSWQ is a 16-item measure whose developers (Meyer, Miller, Metzger & Borkovec, 1990) report test-retest reliability Pearson r_s of .75 at 2 weeks and .74-.93 at 4 weeks, and high internal consistency (alpha = .95). College students who met all Diagnostic and Statistical Manual of Mental Disorders-III-Revised (DSM-III-R; American Psychiatric Association, 1987) criteria for Generalized Anxiety Disorder scored higher on the PSWQ than did those who met criteria for Post Traumatic Stress Disorder or those with no GAD symptoms (Meyer et al., 1990).

Davey (1993) replicated the internal consistency findings (alpha = .94) as did Brown, Antony, and Barlow (1992) (alpha = .93) using a very large (n = 436) clinical sample. The Brown et al. (1992) study also supported the convergent and criterion validity of the PSWQ in that subjects diagnosed with Generalized Anxiety Disorder scored higher than both other anxiety patients and normal controls,

and that those subjects' PSWQ scores correlated significantly with an independent measure of tension.

Social Phobia and Anxiety Inventory (SPAI)

The SPAI (Turner, Beidel, Dancu & Stanley, 1989) contains 45 items related to somatic symptoms, cognitions, anxiety, and escape or avoidance behaviors associated with social phobia. Subjects are asked to assess symptom frequency on a 7-point scale ranging from 1 (never) to 7 (always).

The SPAI has two subscales, one for agoraphobia and one for social phobia, and a Total or Difference score obtained by subtracting the Agoraphobia Subscale score from the Social Phobia Subscale score. This procedure has the Agoraphobia Subscale score serving "as a suppressor variable" allowing "finer differentiation between these two conditions." (Turner et al., 1989, p. 37). However, see Herbert, Bellack and Hope (1991) for cautions about when to use the Difference score. They suggest, for example, that in subjects with symptoms of both agoraphobia and social phobia, using the Difference score may produce false negatives.

In a series of psychometric studies, the SPAI has been shown capable of discriminating the socially anxious from other anxiety groups diagnosed by clinical interviews (Beidel, Turner, Stanley & Dancu, 1989; Turner et al., (1989). It has a 2-week test-retest reliability of .86

(Turner et al., 1989). It has shown good convergent validity with a number of other measures of social anxiety (Herbert, Bellack & Hope, 1991).

CHAPTER 3

RESULTS

The means, standard deviations, and frequency distributions for the BAI, PSWQ, FQ, and for the social phobia and agoraphobia subscales in the FQ and SPAI were computed for the complete samples and also separately for men compared to women; Blacks[1] compared to non-Hispanic Caucasians; and for the lowest income group compared to the other income groups combined (see Tables 3, 6, 7 & 8). The planned number of subgroup comparisons was held to these three in order to keep the studywise Type I error rate low.

Because the church-recruited Hispanic subjects differed from the others in recruiting method, incentive, and administration procedures, the data were also analyzed excluding their responses. Although excluding these subjects worsened the mean scores by about 1 point, the differences were not statistically significant. Their data therefore are included in the results reported.

To allow for possible non-normal distributions, both independent t tests and Wilcoxon nonparametric analyses were computed on the subgroup means using SAS software. Since

[1] Although African-American may be the currently preferred phrasing in the United States, the term Blacks is used here to take into account the presence of a few non-Americans in the samples.

the pattern of results was the same for both sets of analyses, only *t* test results are reported.

For *t* tests SAS performs an *F* test to assess homogeneity of variance between groups and provides an estimated *t* to be used where the *F* test indicates a high probability that the variances are unequal. Where variance was assessed as unequal, the estimated *t* is reported.

In all cases, higher scores indicate higher symptom levels.

Fear Questionnaire

Results of the FQ reported here (Tables 3 & 9) do not include Items 1, 17, or 23--the write-in items. Out of a possible maximum of 80, the range of responses obtained on the Agoraphobia subscale was 0-36 and for the Social Phobia subscale, 0-32. Out of a possible maximum score of 120, the range of Total Phobia scores was 0-86. Comparisons to the equivalent data from the previous two normative studies for the FQ are presented in Table 4.

The FQ norms obtained in this study fall between those obtained in the two previous studies, and within one standard deviation (SD) of both. For the Agoraphobia and Social Phobia subscales, the means obtained here are closer by about one-half SD to the Mizes and Crawford (1986) results, but the Total Phobia score falls closer to the Trull and Hillerbrand (1990) results.

There were no significant gender or race differences,

nor any differences between the lowest income group and the top 80% (Table 3).

Item-corrected/total correlations for the five Agoraphobia subscale questions ranged from .49-.72 and for the five Social Phobia subscale items .46-.53. This higher internal consistency finding for the Agoraphobia subscale was also reported by Trull and Hillerbrand (1990). Item-corrected/total correlations for the 20 items included in the Total score ranged from .30-.62.

Beck Anxiety Inventory

The BAI results are presented in Tables 6 and 10. There were no significant differences between any of the subgroups. The range of scores was 0-43 out of a possible maximum of 60.

BAI item-corrected/total correlations ranged from .31-.72. Borden, Peterson and Jackson (1991) reported item-total correlations ranging from .41-.70 in a nonclinical sample. Items 1 ("Numbness or tingling"), 3 ("Wobbliness in the legs"), and 19 ("Faint") received no endorsements at the most severe level.

Penn State Worry Questionnaire

Data for the PSWQ are presented in Tables 7 and 10. The range of total scores was 18-76 out of a possible 16-80. There were no significant differences between the subgroups.

Item-corrected/total correlations for the five PSWQ reverse-scored items ranged from .23 to .39. The

correlations for the 11 normally scored PSWQ items ranged from .57-.72. Brown et al. (1992) reported item-total correlations ranging between .31 and .85 (M = .65) for a sample of anxiety disordered subjects. Item 1 ("If I don't have enough time to do everything, I don't worry about it."), a reverse-scored item, represented the lowest and only unsatisfactory correlation. All items were endorsed across the full range of possible answers.

Social Phobia and Anxiety Inventory

The means, standard deviations, *t* values and probability for the Social Phobia and Anxiety Inventory results are presented in Table 8. Percentile scores are presented in Table 11. The Social Phobia subscale scores ranged from 0-175 out of a possible maximum of 192. The Agoraphobia subscale scores ranged from 0-76 out of a possible maximum of 78.

When the subgroups were compared, there were no significant differences between the sexes on the Social Phobia subscale or Difference score. However, women scored significantly worse than men on the Agoraphobia subscale (Table 8). In addition, the Social Phobia subscale scores and the scores for the lowest income group were significantly worse than the scores for the combined higher income groups. The score differences between Caucasians and Blacks were also significant for the Difference score and for the Social Phobia subscale, with Caucasians scoring

worse.

Item-corrected/total correlations for the 32 items of the Social Phobia subscale ranged from .48 to .84. Eight items (Items 9, 10, 12, 13, 14, 16, 19, and 22); had correlations over .70 (range .74-.84).

All items were endorsed across the full range of options. However, the modal response to item 5 ("I feel anxious when making a speech in front of an audience") was the most extreme response possible. Nearly 26% of subjects endorsed this item at that level. While some other items, notably discussing intimate feelings, handling embarrassing moments, and being criticized/rejected (Items 14, 15, and 23) were endorsed at the most extreme level by 10-17% of subjects, for those three items the extreme responders were offset by larger numbers who endorsed it at lower levels, so that the frequency distribution remained skewed toward the lower levels of endorsement.

Item-corrected/total correlations for the Agoraphobia subscale ranged from .50 to .76.

CHAPTER 4

DISCUSSION

This study was designed to provide normative data on four psychometrically sound measures of anxiety frequently used in psychotherapy outcome research, particularly for social phobia, agoraphobia, and generalized anxiety disorder. Availability of community norms on these measures would facilitate assessment of the clinical significance of posttreatment improvement as distinguished from statistical significance.

Internal Consistency of the Measures

Item analyses of the questionnaires used in this study confirmed that all have high internal consistency. The SPAI may be longer than necessary, in that eight items had item-corrected/total correlations exceeding .70. It also appears that the internal consistency of the PSWQ, while high, may be moderated by the reverse-scored items whose item-corrected/total correlations were uniformly lower than the normally scored items. Possibly some subjects failed to use the reverse scoring key.

Subgroup Comparisons

Three subgroup comparisons were conducted for norming purposes because there is evidence from epidemiological

studies that there are significant gender, race, and socio-economic differences in prevalence of anxiety disorders. In particular, Boyd et al. (1990) reported that, based on ECA data, women were significantly more likely than men to report a one-month prevalence of agoraphobia or simple phobia, but not social phobia. Similarly, at some ECA sites, the lowest socioeconomic quartile was significantly more likely to have a phobia compared to the highest socioeconomic quartile. Age and gender are also associated with variations in symptom prevalence across a broad range of disorders (e.g., Bebbington, Dean, Der, Hurry & Tennant, 1991; Himmelfarb, 1984; Pollard & Henderson, 1988; Regier et al., 1990).

Previous findings regarding gender differences in responses to the FQ have been mixed. No gender differences were found on FQ responses in this study. Nor were there any gender differences in responses to the PSWQ or the BAI. Brown et al. (1992) also failed to find gender differences in PSWQ scores. Borden et al., (1991), however, reported higher BAI scores for women than men in a sample of undergraduates. Only the SPAI results conformed to the epidemiological pattern, with women scoring significantly worse than men on the SPAI Agoraphobia subscale but not the Social Phobia subscale.

With regard to race, only one racial difference was found here. On the SPAI, Blacks scored significantly better

than Caucasians on both the Social Phobia Subscale and the Difference score. Boyd et al. (1990) reported that phobias, including agoraphobia and social phobia, were "slightly more common among Blacks" (p. 318). Although this proved true only for the St. Louis ECA site, with no differences found in Baltimore, New Haven, Durham, or Los Angeles, it is nonetheless somewhat unexpected that Blacks in this study reported fewer phobic symptoms on this measure than did Caucasians. Examination of the distribution of incomes by race shows that for this measure, only two of the 53 subjects in the lowest income category for the SPAI were Black. The mean income for Blacks in the SPAI sample, $38,000, exceeded that of Caucasians by $4,000. Race, therefore, may have been confounded by income, since the lowest income group also scored significantly worse than higher income subjects on the Social Phobia subscale and Difference score.

There were no differences between the two income groups, however, on SPAI Agoraphobia Subscale scores or on any of the other measures. Since the comparison here was between the lowest quintile of income and all other income groups, whereas the ECA study compared the lowest socioeconomic quartile to the highest socioeconomic quartile, with socioeconomic status being based on income, education, and occupation, perhaps these differences account for the discrepancy between the two sets of findings.

Issues Affecting Generalizability

Moreover, failure to find significant subgroup differences here does not rule out the existence of such differences. Although the subjects here resemble the national census profile with regards to age, gender, race, and income, because they represent a convenience rather than a probability sample, the degree to which they differ from the general population is unknown. Given the strong psychometric properties of the instruments in this study, it is likely that demographic differences may account for the varied findings. Age, in particular student vs. post-collegiate contrasts, and community size may be particularly likely to interact and influence reported fear levels. Given the intersite differences in the ECA data that have not been satisfactorily explained (Boyd et al., 1990), further research explicitly addressing these factors seems warranted.

It should also be noted that, relative to a sample excluding members of the clinical population, the mean of this sample may be skewed towards dysfunctionality, lowering the point to which normality will be anchored and resulting in a less stringent comparison criterion. For example, three of the subjects included in these samples wrote notes indicating that they suffered from diagnosed mental disorders. On the other hand, for all but 25 of the subjects, the sampling sites (shopping malls and a church)

are ones avoided by many agoraphobics, so scores for this disorder may be more stringent than for the others. Likewise, socially anxious subjects might have been more likely to avoid the initial screening contact.

Discrepant Fear Questionnaire Norms

With regards to the FQ, the goal in obtaining a third set of normative data had been to confirm one or the other of the discrepant results from the two previous studies. Confirming the Fargo results might have suggested that the previous discrepancy was due to differences in administration. That conclusion might also have been bolstered if inclusion/exclusion in this study of the data from the subset of church-recruited Hispanics who received oral translation of the questions had altered the results. It did not. Moreover, neither of the previous results was replicated; instead a third estimate was obtained, one that falls well within the 90% confidence intervals for both the previously obtained means. This suggests that the differences may be the result of normal sampling error in a population with large variance. On the other hand, differences in demographic profiles or response rates could be factors. Mizes and Crawford, obtained initial agreements to participate by telephone and followed up by mailing the questionnaires. They report a response rate of 88%, with 68% of returned questionnaires being usable. Trull and Hillerbrand randomly contacted subjects by phone and

administered the FQ verbally during that same contact. They report response rates of 72%. In this study, subjects were approached via a mall intercept and screened to match a demographic profile of the general population. Only subjects matching the profile were asked to participate with any refusers automatically being replaced by a someone similar with respect to the screening variables. Comparative demographic profiles of the three samples are presented in Table 5. Compared to the samples from this study, the other samples included more women, fewer minorities, and, on average, older subjects. For all the measures in this study, additional norms from a variety of additional sites/communities would be desirable to increase confidence in the generalizability of the data.

Sample Calculations of Clinical Significance

Until such additional data are collected, the results presented here enable researchers and clinicians to calculate any of the three alternatives presented by Jacobson et al. (1984) or straightforward comparisons of clients' rankings relative to this sample. For example, using Jacobson et al.'s (1984) definition 2 for clinical significance (posttreatment score should fall within two standard deviations of the normal population), on the PSWQ subjects should score no higher than 64.8.

Using their definition 1 (the posttreatment score should fall two standard deviations from the dysfunctional

population's mean) does not directly require the data provided here. Instead, operationalizing the dysfunctional population's mean and standard deviation as the pretreatment mean and standard deviation for the dysfunctional group, one would calculate a posttreatment score two standard deviations higher than that as being clinically significant. However, that criterion should then be compared to the normative data here. For example, based on the dysfunctional group data provided by Brown et. al. (1992), using this method, one would set a score of 48.93 or less as necessary to be clinically significant following treatment for generalized anxiety disorder. Since that is still above the mean obtained here (more dysfunctional), yet better than the most dysfunctional quarter of the normative group, it probably would not be considered excessively stringent or lenient.

For definition 3 (the posttreatment score should be more likely to have come from the functional than the dysfunctional population), again the dysfunctional population mean could be assumed to be the pretreatment group mean. Using the Beck and Steer (1991) BAI data, since the two groups' variances are unequal, the cutoff score would be calculated according to the formula provided by Jacobson and Truax (1990). In situations where the variances were equal, one would add the mean obtained here (6.6) to the pretreatment dysfunctional mean (e.g., 23.0)

and divide by 2 to produce a cutoff score of 14.8. Scores below this point would be necessary to claim clinically significant improvement, but note that this cutoff score is at the 80-90th percentile of the normative group. Since no effort was made to screen for dysfunction in the normative sample, a lower percentile score might be a more appropriate target. In fact, instead of using any of the Jacobson et al. (1984) methods, one could simply select the median/50th percentile as the clinically significant cutoff. In that case, based on the SPAI scores presented here, posttreatment scores of 63.4 or better would indicate clinically significant improvement on the Social Phobia subscale. In all the examples cited above, it is assumed that the degree of change has been assessed as reliable (Jacobson & Revenstorf, 1988).

It should be noted that the mean exceeds the median for all the measures, which is to be expected for symptom measures, but which also highlights the fact that even after defining normal as average, there are choices to be made in defining average. It should also be noted that, in all these examples of computing clinical significance, the comparison norm is the score obtained by the overall sample. Presentation of subgroup data here is not intended as a blanket recommendation for use of subgroup norms. Where significant subgroup differences were found, a decision whether to use subgroup norms, and which subgroup comparison

to make, would have to be made. When and whether to use subgroup norms is a complex decision which each clinician and researcher must make based on the specifics of their population characteristics and context. For example, the norms suggest that social phobia may be present to a greater degree in the lowest income groups. Taking into consideration the possibility that social anxiety might hamper job searching or job performance relative to the overall population, improvement goals pegged to the overall population could have significantly greater life impact than goals linked to the subgroup level.

APPENDIX

TABLE 1

DEMOGRAPHIC QUOTAS

	U.S. Census	FQ Sample	BAI Sample	PSWQ Sample	SPAI Sample
	%	%	%	%	%
GENDER					
Men	49.0	48.0	47.3	46.9	49.6
Women	51.0	52.0	52.7	53.1	50.4
RACE					
Caucasian	84.3	85.1	84.4	84.9	84.5
(Hispanic)	(9.0)	(10.3)	(10.9)	(11.0)	(8.4)
Blacks	12.3	11.6	12.6	11.8	12.2
Other	3.4	3.3	2.9	3.3	3.3
INCOME					
<$16K	20	20.1	20.0	20.7	17.6
AGE					
18-29	32.1	29.4	31.8	31.8	35.7
30-44	37.6	42.1	40.1	40.9	37.4
45-64	30.3	28.5	28.1	27.3	30.3

TABLE 2

DEMOGRAPHIC DATA

	FQ		BAI		PSWQ		SPAI	
	Data Used	Not Used	Data Used	Not Used	Data Used	Not Used	Data Used	Not Used
AGE								
Below 30	29%	42%	32%	40%	32%	46%	36%	27%
30-44	42%	38%	40%	44%	41%	41%	37%	36%
45-64	29%	21%	28%	16%	27%	14%	27%	36%
RACE								
Caucasian	85%	76%	84%	83%	85%	77%	85%	74%
(Hispanic)	(10%)	(12%)	(11%)	(4%)	(11%)	(5%)	(8%)	(13%)
Black	12%	20%	13%	13%	12%	18%	12%	13%
Other	3%	4%	3%	4%	3%	5%	3%	13%
EDUCATION								
< HS	4%	5%	4%	5%	4%	5%	2%	-
High School Diploma	45%	64%	46%	57%	45%	60%	46%	70%
College Degree	32%	27%	32%	33%	32%	30%	33%	15%
Graduate Degree	14%	-	13%	5%	13%	5%	13%	15%
Unreported	5%	-	5%	5%	5%	-	6%	-
GENDER								
Male	48%	40%	47%	52%	47%	50%	50%	39%
Female	52%	60%	53%	48%	53%	50%	50%	61%

TABLE 2 (continued)

	FQ		BAI		PSWQ		SPAI	
	Data Used	Not Used	Data Used	Not Used	Data Used	Not Used	Data Used	Not Used
MARITAL STATUS								
Married	39%	42%	39%	39%	40%	36%	36%	30%
Divorced	14%	8%	14%	9%	13%	9%	11%	14%
Never Married	41%	46%	41%	48%	40%	50%	42%	48%
Widowed/ Widower	4%	4%	3%	4%	4%	5%	5%	-
Separated	3%	-	4%	-	3%	-	5%	9%
EMPLOYMENT STATUS								
Fulltime	48%	33%	47%	44%	48%	36%	46%	48%
Parttime	9%	13%	9%	13%	9%	14%	12%	4%
Unemployed	14%	21%	13%	26%	13%	32%	13%	22%
Student	11%	13%	11%	9%	11%	95%	14%	4%
Retired	11%	13%	12%	9%	12%	9%	11%	13%
Self-employed	7%	-	7%	8%	7%	8%	5%	9%
INCOME								
< $16K	21%	22%	21%	25%	21%	29%	19%	20%
$16K -$27.9K	18%	57%	19%	45%	19%	47%	20%	40%
$28K- $40.9K	22%	17%	22%	20%	22%	14%	24%	20%
$41K- $60K	15%	-	15%	-	15%	-	14%	10%
Over $60K	24%	4%	23%	10%	23%	10%	23%	10%

TABLE 3

FEAR QUESTIONNAIRE MEANS

	Overall (n = 240)	Men vs. Women		Blacks vs. Caucasians		< $16K vs. $16K & Over	
		(n = 116)	(n = 124)	(n = 28)	(n = 179)	(n = 47)	(n = 193)
Social Phobia Subscale							
Mean	10.8	10.6	11.0	10.8	11.0	10.5	10.5
SD	6.8	7.0	6.6	8.3	6.2	7.2	6.7
t	--	1.30		1.70		.20	
p	--	.20		.09		.84	
Agoraphobia Subscale							
Mean	6.5	5.8	7.0	9.5	6.0	6.2	6.4
SD	7.2	6.9	7.4	10.3	6.4	7.1	7.1
t	--	.44		.11		.35	
p	--	.66		.91		.73	
TOTAL PHOBIA							
Mean	37.8	38.3	37.4	43.5	37.7	36.0	38.1
SD	22.4	22.3	22.5	28.2	20.7	22.2	22.4
t	--	.31		1.0		.59	
p	--	.76		.31		.56	

TABLE 4

FEAR QUESTIONNAIRE -- 3-SITE COMPARISON: RESULTS

	T & H		M & C		WDC	
	M (n = 48)	F (n = 63)	M (n = 109)	F (n = 107)	M (n = 116)	F (n = 124)
Social Phobia Subscale						
Mean	14.2	16.3	8.9	8.8	10.6	11.0
SD	(4.3)	(6.2)	(5.5)	(6.1)	(7.0)	(6.6)
Agoraphobia Subscale						
Mean	7.9	14.8	4.9	5.0	5.8	7.0
SD	(7.1)	(8.5)	(5.1)	(6.0)	(6.9)	(7.4)
Blood/Injury Subscale						
Mean	10.9	15.0	9.1	7.3	11.9	10.8
SD	(7.1)	(6.8)	(6.5)	(6.0)	(7.9)	(8.3)
Anxiety Subscale						
Mean	NR	NR	9.0	6.9	10.0	8.6
SD	NR	NR	(6.7)	(6.2)	(7.2)	(6.3)

TABLE 4 (continued)

	T & H		M & C		WDC	
	M	F	M	F	M	F
TOTAL PHOBIA						
Mean	33.0	46.1	22.8	21.2	38.3	37.4
SD	(13.1)	(16.2)	(14.1)	(14.5)	(22.3)	(22.5)

WDC = Washington, DC, 1994
T&H = Trull & Hillerbrand, 1990
M&C = Mizes & Crawford, 1986

SD = Standard Deviation

M = Males
F = Females

NR = Not reported

Range of Possible Scores: Agoraphobia subscale: 0-40
Social phobia subscale: 0-40
Blood/Injury subscale: 0-40
Depression/Anxiety subscale: 0-40
TOTAL PHOBIA: 0-120

TABLE 5

FEAR QUESTIONNAIRE--3-SITE COMPARISON: DEMOGRAPHICS

	U.S. Census %	WDC Sample %	M & C Sample %	T & H Sample %
GENDER				
Men	49.0	48.0	42.4	43.0
Women	51.0	52.0	57.6	57.0
RACE				
Caucasian	84.3	85.1	97.1	NR
(Hispanic)	(9.0)	(10.3)	(1.7)	NR
Blacks	12.3	11.6	0.6	NR
Other	3.4	3.3	0.6	NR
INCOME				
<$16K	20.0	20.1	>20.9[2]	NR
AGE				
18-29	32.1	29.4		
30-44	37.6	42.1		
45-64	30.3	28.5		
Mean	37.0	41.5	42.6	

[2] Mizes and Crawford reported 20.9% at or below $15,000, whereas the income cutoff for the other two studies was $16,000.

TABLE 5 (continued)

	U.S. Census	WDC Sample	M & C Sample	T & H Sample
Age Range	--	18-64	20-88	NR

NR = Not Reported
WDC = Washington, DC, 1994
T&H = Trull & Hillerbrand, 1990
M&C = Mizes & Crawford, 1986

TABLE 6

BECK ANXIETY INVENTORY

	Overall (n = 247)	Men vs. Women (n = 113)	(n = 126)	Blacks vs. Caucasians (n = 30)	(n = 175)	< $16K vs. $16K & Over (n = 48)	(n = 192)
Mean	6.6	6.3	6.7	9.2	6.3	6.6	6.6
SD	8.1	7.2	8.7	9.2	7.8	8.4	7.0
t	--	.43		1.9		<.01	
p	--	.66		.06		1.0	

SD = Standard Deviation

Maximum score possible = 63

TABLE 7

MEAN PENN STATE WORRY QUESTIONNAIRE SCORES

	Overall	Men vs. Women		Blacks vs. Caucasians		<$16K vs. $16K & Over	
	(n = 245)	(n = 115)	(n = 130)	(n = 29)	(n = 181)	(n = 48)	(n = 194)
Mean	36.5	41.4	42.9	44.7	41.3	42.8	41.9
SD	9.9	11.7	10.9	12.3	11.5	11.0	11.4
t	--	1.07		1.52		.48	
p	--	.28		.13		.63	

Range of possible scores: 0-80

TABLE 8

MEAN SOCIAL PHOBIA AND ANXIETY INVENTORY SCORES

	Overall	Men vs. Women		Blacks vs. Caucasians		<$16K vs. $16K & Over	
	(n = 228)	(n = 118)	(n = 120)	(n = 29)	(n = 181)	(n = 42)	(n = 190)
Social Phobia Subscale							
Mean	65.6	62.6	68.6	50.4	68.1	77.9	63.0
SD	34.2	31.0	37.0	29.5	34.7	39.1	32.6
t		1.4		2.6	2.3		
p		.17		.01		.01	
Agoraphobia Subscale							
Mean	18.0	14.9	21.1	17.8	17.3	20.3	17.5
SD	14.4	10.6	16.8	16.3	13.2	16.7	13.8
t		3.4		.2		1.0	
p		<.01		.84		.26	

TABLE 8 (continued)

SOCIAL PHOBIA AND ANXIETY INVENTORY SCORES

	Overall	Men vs. Women		Blacks vs. Caucasians		<$16K vs. $16K & Over	
	(n = 228)	(n = 118)	(n = 120)	(n = 29)	(n = 181)	(n = 42)	(n = 190)
Difference Score							
Mean	47.6	47.7	47.5	32.6	50.8	57.6	45.5
SD	28.9	27.5	30.3	21.5	29.1	31.5	27.9
t	--	.1		3.2		2.3	
p		.96		>.01		.01	

Range of Possible Scores:

Social Phobia Subscale: 0-192

Agoraphobia Subscale: 0-78

TABLE 9

FEAR QUESTIONNAIRE - CUMULATIVE RANKINGS

FROM OVERALL SAMPLE

Percentile	Subscales				
	AG	SP	BI	A/D	TOT
50th	4.4	9.2	9.5	7.0	33.5
60th	5.0	11.0	11.6	9.2	38.0
70th	7.0	13.2	13.4	11.2	
45.3					
75th	8.2	14.6	15.2	12.3	
48.9					
80th	9.7	16.0	17.0	13.8	
53.7					
90th	16.0	20.5	22.3	18.8	
65.0					

AG = Agoraphobia subscale

SP = Social Phobia subscale

BI = Blood Injury subscale

A/D = Anxiety/Depression subscale

TOT = Total phobia score

TABLE 10

CUMULATIVE PERCENTILE RANKINGS FOR THE

BECK ANXIETY INVENTORY AND PENN STATE WORRY QUESTIONNAIRE

	BAI	PSWQ
	Overall Sample Score	Overall Sample Score
50th percentile	3.3	40.9
60th percentile	4.9	44.1
70th percentile	7.1	47.7
75th percentile	8.4	48.9
80th percentile	10.4	50.6
90th percentile	17.4	57.1

TABLE 11

SOCIAL PHOBIA AND ANXIETY INVENTORY

OVERALL CUMULATIVE PERCENTILE RANKINGS

Percentile	Overall Sample Scores		
	SP	AG	TOT
50th	63.4	14.4	46.8
60th	72.7	17.5	52.8
70th	82.8	20.9	62.4
75th	90.8	22.9	67.0
80th	96.1	25.7	70.4
90th	113.9	37.1	87.0

AG = Agoraphobia Subscale score

SP = Social Phobia Subscale score

TOT = Difference score

TABLE 12

SOCIAL PHOBIA AND ANXIETY INVENTORY

SUBGROUP CUMULATIVE PERCENTILE RANKINGS

Percen-	Blacks			Whites			Higher Incomes			Lowest Income		
tile	SP	AG	TOT	SP	AG	TOT	SP	AG	TOT	SP	AG	TOT
50th	47.8	15.5	30.2	68.2	13.1	49.2	61.7	13.8	46.2	74.3	15.0	56.4
60th	55.4	18.4	32.2	75.9	16.4	56.7	70.6	17.0	51.9	80.9	21.2	62.3
70th	60.4	21.9	37.1	85.2	20.1	65.0	80.7	20.7	60.1	98.2	22.9	68.0
75th	69.1	24.8	42.5	92.0	21.9	68.0	88.1	22.0	65.9	100.4	25.0	68.0
80th	70.6	27.4	46.7	97.2	24.4	73.2	93.8	25.0.	69.6	112.6	31.7	82.3
90th	88.6	35.1	64.1	115.4	33.9	88.0	105.4	35.2	83.9	138.6	49.2	103.8

AG = Agoraphobia Subscale score

SP = Social Phobia Subscale score

TOT = Difference score

REFERENCES

American Psychiatric Association. (1987). Diagnostic and statistical manual of mental disorders 3rd ed.-revised. Washington, DC: Author.

Arrindell, W. A., & Buikhuisen, M. (1992). Dissimulation and the sex difference is self-assessed fears: A brief note. Behaviour Research and Therapy, 30, 307-311.

Baer, D. M. (1988). If you know why you're changing a behavior, you'll know when you've changed it enough. Behavioral Assessment, 10, 219-223.

Bebbington, P. E., Dean, C., Der, G., Hurry, J., & Tennant, C. (1991). Gender, parity and the prevalence of minor affective disorders. British Journal of Psychiatry, 158, 40-45.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. Journal of Consulting and Clinical Psychology, 56, 893-897.

Beck, A. T., & Steer, R. A. (1991). Relationship between the Beck Anxiety Inventory and the Hamilton Anxiety Rating Scale with anxious outpatients. Journal of Anxiety Disorders, 5, 213-223.

Beidel, D. C., Turner, S. M., & Cooley, M. R. (1993). Assessing reliable and clinically significant change in social phobia: Validity of the Social Phobia and Anxiety Inventory. Behaviour Therapy and Research, 31, 331-337.

Beidel, D. C., Turner, S. M., Stanley, M. A., & Dancu, C. V. (1989). The Social Phobia and Anxiety Inventory: Concurrent and external validity. Behavior Therapy, 20, 417-427.

Blanchard, E. B., & Schwarz, S. P. (1988). Clinically significant changes in behavioral medicine. Behavioral Assessment, 10, 171-188.

Borden, J. W., Peterson, D. R., & Jackson, E. A. (1991).

The Beck Anxiety Inventory in nonclinical samples: Initial psychometric properties. Journal of Psychopathology and Behavioral Assessment, 13, 345-356.

Borkovec, T. D., & Mathews, A. M. (1988). Treatment of nonphobic anxiety disorders: A comparison of nondirective, cognitive, and coping desensitization therapy. Journal of Consulting and Clinical Psychology, 56, 877-884.

Boyd, J. H., Rae, D. S., Thompson, J. W., Burns, B. J., Bourdon, K., Locke, B. Z., & Regier, D. A. (1990). Phobia: Prevalence and risk factors. Social Psychiatry and Psychiatric Epidemiology, 25, 314-323.

Brown, T. A., Antony, M. M., & Barlow, D. H. (1992). Psychometric properties of the Penn State Worry Questionnaire in a clinical anxiety disorders sample. Behaviour Research and Therapy, 30, 33-37.

Chambless, D. L., & Gillis, M. M. (1993). Cognitive therapy of anxiety disorders. Journal of Consulting and Clinical Psychology, 61, 248-260.

Cox, B. J., Swinson, R. P., & Fergus, K. D. (1993). Changes in fear versus avoidance ratings with behavioral treatments for agoraphobia. Behavior Therapy, 24, 619-624.

Cox, B. J., Swinson, R. P., & Shaw, B. F. (1991). Value of the Fear Questionnaire in differentiating Agoraphobia and Social Phobia. British Journal of Psychiatry, 159, 842-845.

Davey, G. C. (1993). A comparison of three worry questionnaires. Behaviour Research and Therapy, 31, 51-56.

Fydrich, T., Dowdall, D., Chambless, D. L. (1992). Reliability and validity of the Beck Anxiety Inventory. Journal of Anxiety Disorders, 6, 55-61.

Heimberg, R. G. (1989). Cognitive and behavioral treatments for social phobia: A critical analysis. Clinical Psychology Review, 9, 107-128.

Herbert, J. D., Bellack, A. S., & Hope, D. A. (1991). Concurrent validity of the Social Phobia and Anxiety Inventory. Journal of Psychopathology and Behavioral Assessment, 13, 357-368.

Himmelfarb, S. (1984). Age and sex differences in the mental

health of older persons. Journal of Consulting and Clinical Psychology, 52, 844-856.

Hollon, S. D., & Flick, S. N. (1988). On the meaning and methods of clinical significance. Behavioral Assessment, 10, 197-206.

Hugdahl, K., & Ost, L-G. (1981). On the difference between statistical and clinical significance. Behavioral Assessment, 3, 289-295.

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. Behavior Therapy, 15, 336-352.

Jacobson, N. S., & Revenstorf, D. (1988). Statistics for assessing the clinical significance of psychotherapy techniques: Issues, problems, and new developments. Behavioral Assessment, 10, 133-145.

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.

Kazdin, A. E., & Bass, D. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. Journal of Consulting and Clinical Psychology, 57, 138-147.

Kazdin, A. E., & Wilson, G. T. (1978). Criteria for evaluating psychotherapy. Archives of General Psychiatry, 35, 407-416.

Kendall, P. C., & Grove, W. M. (1988). Normative comparisons in therapy outcome. Behavioral Assessment, 10, 147-158.

Lelliott, P., McNamee, G., & Marks, I. (1991). Features of Agora-, Social, and related phobias and validation of the diagnoses. Journal of Anxiety Disorders, 5, 313-322.

Marks, I. M., & Mathews, A. M. (1979). Brief standard self-rating for phobic patients. Behavior Research and Therapy, 17, 263-267.

Mavissakalian, M. (1986). The Fear Questionnaire: A validity study. Behaviour Research and Therapy, 24, 83-85.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. Behavior Research and Therapy, 28, 487-495.

Mizes, J. S., & Crawford, J. (1986). Normative values on the Marks and Mathews fear questionnaire: A comparison as a function of age and sex. Journal of Psychopathology and Behavioral Assessment, 8, 253-262.

Moylan, A. & Oei, T. P. S. (1992). Is the Fear Questionnaire FQ a useful instrument for patients with anxiety disorders? Behaviour Change, 9, 38-49.

Nietzel, M. T., & Trull, T. J. (1988). Meta-analytic approaches to social comparisons: A method for measuring clinical significance. Behavioral Assessment, 10, 159-169.

Oei, T. P. S., Moylan, A., & Evans, L. (1991). Validity and clinical utility of the Fear Questionnaire for anxiety-disorder patients. Psychological Assessment, 3, 391-397.

Pollard, C. A., & Henderson, J. G. (1988). Four types of social phobia in a community sample. Journal of Nervous and Mental Disease, 176, 440-445.

Regier, D. A., Narrow, W. E., & Rae, D. S. (1990). The epidemiology of anxiety disorders: The Epidemiologic Catchment Area ECA experience. Journal of Psychiatric Research, 24, 3-14.

Reich, J. (1986). The epidemiology of anxiety. The Journal of Nervous and Mental Disease, 174, 129-136.

Robins, L. N., Helzer, J. E., Weissman, M. M., Orvaschel, H., Gruenberg, E., Burke, J. D., & Regier, D. A. (1984). Lifetime prevalence of specific psychiatric disorders in three sites. Archives of General Psychiatry, 41, 949-958.

Saunders, S. M., Howard, K. I., & Newman, F. L. (1988). Evaluating the clinical significance of treatment effects: Norms and normality. Behavioral Assessment, 10, 207-218.

Turner, S. M., Beidel, D. C., Dancu, C. V., & Stanley, M. A. (1989). An empirically derived inventory to measure social fears and anxiety: The Social Phobia and Anxiety Inventory. Psychological Assessment: A

Journal of Consulting and Clinical Psychology, 1, 35-40.

Turner, S. M., Beidel, D. C., Long, P. J., Turner, M. W., & Townsley, R. M. (1993). A composite measure to determine the functional status of treated social phobics: The Social Phobia Endstate Functioning Index. Behavior Therapy, 24, 265-275.

U.S. Department of Commerce, Bureau of the Census. (1990). Statistical Abstract of the United States 1990: The National Data Book 110th ed.. Washington, D.C.: Author.

Trull, T. J. & Hillerbrand, E. (1990). Psychometric properties and factor structure of the Fear Questionnaire Phobia Subscale items in two normative samples. Journal of Psychopathology and Behavioral Assessment, 12, 285-297.

Wampold, B. E., & Jenson, W. R. (1986). Clinical significance revisited. Behavior Therapy, 17, 302-305.