Independent Evaluator Knowledge of Treatment in a Multicenter Comparative Treatment Study of Panic Disorder

David Roll*[,1], Susan E Ray[2], Sue M Marcus[3,4], Vincent Passarelli[1], Roy Money[5], David H Barlow[6], Scott W Woods[5], M Katherine Shear[7]and Jack M Gorman[3]

[1]Clinical Psychology Doctoral Program, Long Island University-CW Post Campus, Brookville, NY, USA; [2]The Zucker Hillside Hospital, North Shore-Long Island Jewish Health System, Glen Oaks, NY, USA; [3]Department of Psychiatry, Mount Sinai School of Medicine, One Gustave L Levy Place, New York, NY, USA; [4]Departments of Biomathematics, Mt Sinai School of Medicine, USA; [5]Department of Psychiatry, Yale University School of Medicine, Connecticut Mental Health Center, New Haven, CT, USA; [6]Center for Anxiety and Related Disorders & Department of Psychology, Boston University, Boston, MA, USA; [7]Department of Psychiatry, Western Psychiatric Institute and Clinic, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

The purpose of this study was to examine independent evaluators' (IEs) blindness to treatment condition during a Multicenter Comparative Treatment Study of Panic Disorder. IEs were 15 doctoral- and masters-level clinicians in psychology, social work, and medicine. They conducted three post-treatment assessments with each patient. Immediately after each assessment interview, IEs completed a form indicating which of the five possible treatments they believed the patient had received and any specific information that provided IEs with information about a patient's treatment condition. These forms were completed for 170 patients. Analyses were conducted to determine the accuracy of guesses about treatment condition by IEs during post-treatment assessments, the relationship between accuracy of IE guessing and actual treatment assignment, the relationship between accurate guessing and outcome ratings, and contributors to the breaking of the blind. A significant relationship was found between IE guesses and actual treatment at all three assessment points, across individual IEs, treatment sites, and IE professional affiliations. IEs were no more accurate in their guessing about patients taking medication than those receiving behavior therapy. Patients and project staff inadvertently provided information to IEs that enhanced the rates of accurate guessing. Implications of these findings on interpretation of the treatment study are discussed, and recommendations are made for improving blindness procedures.
Neuropsychopharmacology (2004) **29,** 612–618, advance online publication, 31 December 2003; doi:10.1038/sj.npp.1300373

Keywords: clinical research protocols; research methodology; double-blind method; treatment outcome; behavior therapy; pharmacotherapy

INTRODUCTION

An important component in the design of comparative treatment studies concerns the people who evaluate treatment success. It is considered good practice and a defining characteristic of empirically supported treatments for these evaluators to be independent of the treatment process (Kendall, 1998; Klein, 1994) and blind to the treatment condition (Chambless and Hollon, 1998; Seligman, 1995). Seligman (1995) described a defining characteristic of an ideal psychotherapy efficacy study as the use of 'raters and diagnosticians who are blind to which group the patient comes from (p. 965).' As Seligman noted, psychotherapy studies can only be single blind (the evaluator) because therapists and patients always know the treatments. However, placebo-controlled pharmacological studies can be double blind (therapists and patients) or even triple blind (evaluators, therapists, and patients). Truly blind independent evaluators (IEs) as well as double-blind participants are important to the minimization of experimenter bias in clinical trials.

More than three decades ago, Beatty (1972) suggested a simple procedure for assessing the blindness of observers' knowledge of experimental treatments. 'Merely by asking the observer to guess the treatment combinations each subject has received and by comparing the frequency of correct identifications obtained to that expected by chance, one can estimate whether or not the observers were in fact naive (p. 70).' Yet, researchers rarely evaluate or report this aspect of their studies.

A growing number of investigators have reported evaluations of double blindness in pharmacotherapy studies. Fisher and Greenberg (1993) conducted a comprehensive literature search of psychotropic drug trials using double-blind procedures. They reported that in 23 of 26 articles, patients and/or physicians differentiated active from placebo conditions at a rate significantly greater than chance. More recent studies have substantiated the frequent ineffectiveness of double-blind procedures (Double, 1995; Bakker et al, 1999), while others reported that independent evaluators were not able to guess the treatment condition at better than chance rates (Schneier et al, 1998) for pharmacotherapy. None of these articles reported accuracy of guessing treatment conditions by independent evaluators of psychotherapeutic interventions.

Marks et al (1983) reported evaluations of blindness procedures for independent assessors who rated response to both medication (imipramine vs placebo) and behavior therapy (therapist-aided exposure vs relaxation). Results showed that in both cases the guesses of the assessors were no better than would be expected by chance. Carroll et al (1994) were the first to report a thorough analysis of the comparative effectiveness of blinding procedures for independent evaluators in pharmacotherapy and psychological interventions. In a randomized clinical trial on the comparative effectiveness of psychotherapy or a clinical control condition with desipramine or placebo for treatment of cocaine dependence, the authors reported that the rates of correct guesses by independent evaluators were similar for psychotherapy (77%) and pharmacotherapy (75%), and that accuracy in both cases was greater than expected by chance. They also found interactions between treatment conditions and accuracy of guessing. Evaluators guessed most accurately for active psychotherapy when patients were receiving active medication and for the placebo (vs medication) condition when patients were in the control psychotherapy condition. Perhaps the most telling finding of this study was that while subjective outcome ratings were related to the accuracy of independent evaluators' guesses, more behavioral (objective) measures were not.

Basoglu et al (1997) reported a thorough analysis of independent rater blindness in a controlled trial comparing exposure therapies with alprazolam in the treatment of patients with panic disorder and agoraphobia. At the completion of treatment, both Independent Assessors and patients correctly classified participants into drug and placebo conditions at rates greater than chance. The rates of correct classification into exposure (90%) or relaxation (93%) groups for the psychological interventions were also significantly greater than chance. The authors reported that for the drug interventions, side effects contributed to the accuracy of guesses. They also concluded that correct classifications were not related to treatment outcome for psychological or pharmacological interventions.

The present study was designed to examine the extent to which independent evaluators were blind to treatment condition during a Multicenter Comparative Treatment Study of Panic Disorder (MCCTSPD). Specifically, the purposes were to:

1. Determine the accuracy of guesses about treatment condition by (IEs) across three post-treatment assessments.

2. Determine the relationship between accuracy of IE guessing and actual treatment assignment.
3. Determine the relationship between accurate guessing and outcome ratings.
4. Determine variables contributing to the breaking of the blind.

This study also differs from previous studies in that there are five choices of therapeutic conditions rather than the typical two (drug vs placebo), and the present study extends the analysis to three post-treatment assessments rather than the typical one end point.

It is one thing to describe extensive efforts to keep evaluators independent and blind to treatment condition, but quite another to assume that these efforts are always successful. The inclusion of a brief questionnaire asking IEs to indicate which treatment they thought the patient was receiving allowed this determination.

METHODS

A randomized clinical trial of the treatment of panic disorder with no or mild agoraphobia was conducted at four clinical research centers. The study was designed to compare cognitive behavior therapy (CBT); imipramine (IMI); a pill placebo (PLA); CBT + IMI; and CBT + PLA. Randomization was stratified by site and by presence vs absence of DSM-III-R current major depression and was blocked within stratum. In order to improve efficiency (Woods et al, 1998), unequal numbers of patients were randomized to the treatments based on expected pairwise comparison effect sizes (ie, six CBT, six IMI, two PLA, five CBT + IMI, five CBT + PLA per block of 24). IEs were used to conduct the main study assessments and were not told of the unequal randomization.

CBT included psychoeducation, breathing retraining, cognitive restructuring, and interoceptive exposure. Pharmacotherapy (IMI and PLA) was administered in a double-blind, fixed, flexible dose design beginning with 10 mg daily and increasing to at least 200 mg daily or if necessary to 300 mg daily. Medical management included psychoeducation, monitoring of side effects, facilitation of compliance to the medication regimen, and proscription of CBT components. The combined treatment conditions were an integration of the CBT and pharmacotherapy interventions. Further details of the procedures and results can be found in Barlow et al (2000).

The IEs conducted a postacute (PA) evaluation following 12 weeks of treatment and this was used to determine responder status. Responders to all acute treatments and nonresponders to any CBT treatment were continued on the same treatment for an additional 6 months, during which sessions occurred monthly. An identical postmaintenance (PM) independent evaluation was then conducted. Treatment was discontinued, and an identical final follow-up (FF) independent evaluation was conducted 6 months later. Pretreatment assessment measures included the Anxiety Disorders Interview Schedule-Revised (ADIS-R; DiNardo et al, 1988), Structured Clinical Interview for the DSM-III-R-Psychoactive Substance Use Disorders (SCID-PSUD; Spitzer et al, 1988), Panic Disorder Severity Scale (PDSS; Shear et al, 1997), and the Clinical Global

Impressions-Severity (CGI-S; Guy, 1976). Mild agoraphobia was operationalized as a score of 18 or less on the ADIS-R avoidance ratings of situations. All post-treatment independent evaluation measures included those listed above with the exception that the full ADIS-R was replaced by the mini-ADIS-R and with the addition of the Clinical Global Impressions-Improvement (CGI-I; Guy, 1976). Responder status was defined as a CGI-S score of normal, borderline mentally ill, or mildly ill and a CGI-I score of very much improved or much improved.

Each evaluator completed The Independent Evaluator Knowledge of Treatment (IEKNO) following each contact with the patient. The first of the three IEKNO items lists the five treatment conditions and asks which the evaluator thinks the patient is receiving. This was a forced choice in which 'I don't know' was not an option. The second item asks for a certainty rating about the condition assignment rating on a nine-point Likert scale ranging from 'not at all sure' to 'absolutely sure.' The third item is open-ended, requesting any specific information that the IEs believed provided information about the treatment condition.

In total, 326 patients entered this study and IEs completed IEKNO forms for 170/254 (67%) patients who had a PA assessment, 115/174 (66%) assessed at PM, and 82/164 (50%) assessed at FF. The smaller number of ratings at the PA assessment was primarily due to the fact that the IEKNO form was initiated in the second year of the study. The reduced numbers in successive assessments were due to attrition from the study. Some patients who started the study did not reach the post-treatment evaluations.

IEs were 15 doctoral- and masters-level clinicians in psychology, social work, and medicine who underwent extensive training and certification prior to assessing study patients and ongoing supervision during the trial. In attempts to keep IEs blind to treatment assignments patients were instructed in the consent form and usually by both their therapists and their IEs to refrain from mentioning any information that might reveal the treatment condition. However, patients did communicate such information on occasion.

Generally, longitudinal data analysis will lead to incorrect statistical inference if person-specific effects or serial correlation are not considered (Gibbons, 2000). These models can incorporate observations from subjects that may have one or more missing observations. Often the assumption that repeated measurements are equally correlated over time with constant variance is not appropriate for longitudinal psychiatric data. Our analysis was performed using MIXOR software (Hedeker and Gibbons, 1996). These programs can be downloaded from the website http://tigger.cc.uic.edu/~hedeker/long.html. Mixed-effects models were used to model whether the guessing was correct at greater than chance rates over the three assessments (PA, PM, and FF) and as a function of other covariates.

The third item on the IEKNO was an open-ended question requesting any specific information that IEs believed provided clues about the treatment condition. The information from this question was coded independently by two of the authors (VP and RM) into two categories: SLIPS or NO SLIPS. Overall agreement (agreements/agreements + disagreements) between the raters on classification of information into SLIPS and NO SLIPS was 91% with a range across treatment groups of 83% (Placebo) to 94% (IMI). SLIPS included information provided by the patient about medication side effects, practice in behavioral techniques, mention of psychiatrist or therapist, observations of a patient with a therapist or psychiatrist, and unintended information from other staff. NO SLIPS were recorded when IEs made no comment in this part of the questionnaire, wrote that they had no information related to the treatment condition, or the information provided was judged irrelevant to the treatment condition. Data on this measure were recorded for PA only.

RESULTS

IEs could choose one of five possible treatment conditions on the IEKNO questionnaire. If the assumption is made that each category is an equally likely choice for the IEs, 20% of the IE guesses would occur in each category by chance. Table 1 shows IE guesses across treatment groups at the PA assessment. Overall, the true treatment assignment was accurately guessed in 78 of 170 cases (45.9% correct, $\chi^2 = 81.12$, $p = 0.0001$). Using MIXOR, we performed a mixed-effects logistic regression of correct guesses (yes/no) on two dummy variables for PA and PM as well as dummy variables for individual evaluators. We found no significant

Table I IE Classifications of Treatment Groups on the IEKNO at the PA Evaluation

Treatment group	IE classifications of treatment groups					Total patients	% Correct
	CBT/IMI	**CBT/PLA**	**CBT**	**IMI**	**PLA**		
CBT/IMI	**18**	2	6	9	3	38	47.37
CBT/PLA	14	**2**	9	9	2	36	5.56
CBT	4	3	**29**	2	2	40	72.50
IMI	6	1	5	**25**	6	43	58.14
PLA	2	1	1	5	**4**	13	30.77
Total guesses	44	9	50	50	17	170	45.88

Numbers in bold are correct classifications.

effect over time and no significant effect of individual evaluator. We also assessed the association between correct guessing (yes/no) and professional affiliations of the IEs (psychology, social work, or medicine) in a similar way and found no significant difference in correct guessing by IEs in different mental health professions.

Figure 1 shows the rates of correct guessing by IEs at the different treatment sites. The mixed effects analysis of the association between IE correct guesses (yes/no) and treatment site over time shows that the percentage correct guessing by IEs differed significantly for Site 1 vs the others ($p = 0.00625$). On average, Site 1 evaluators guessed about one-third correctly, whereas the other sites guessed about one-half correctly.

The mixed-effects regression of correct (yes/no) on SLIPs over time found the effect of SLIPS on correct guessing to be significant ($p = 0.002$). Since SLIPs were reported only for the PA assessment, we assumed that SLIPs from the first time period carried over to other time periods. When IEs recorded SLIPS, they also rated their confidence in the accuracy of guesses as higher than when SLIPS were not recorded ($t = -0.83$, df $= 168$, $p = 0.0001$). Figures 2 and 3 show the distribution of SLIPS across treatment conditions and the impact of different types of SLIPS on correct classifications of treatment conditions.

Finally, we assessed the relationship between IE correct guessing (yes/no) and SLIPS (yes/no), response to treatment (yes/no), and treatment assignment over time in a multivariate mixed-effects logistic model. Table 2 shows these results. SLIPS are significantly associated with correct response ($p = 0.01$), as is treatment assignment (IMI, $p = 0.0000$; PLA, $p = 0.001$; CBT/IMI, $p = 0.00001$, CBT $= 0.0000$). However, response to treatment is not significantly associated with IE correct guessing in this final multivariate model.
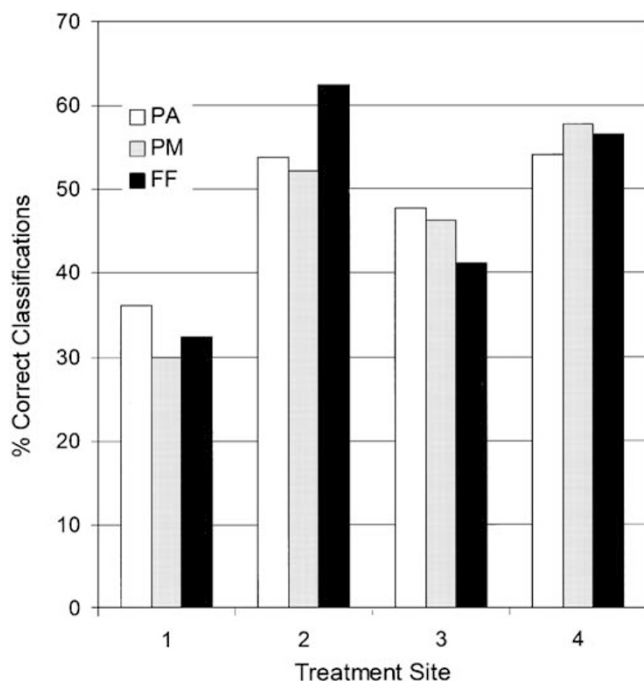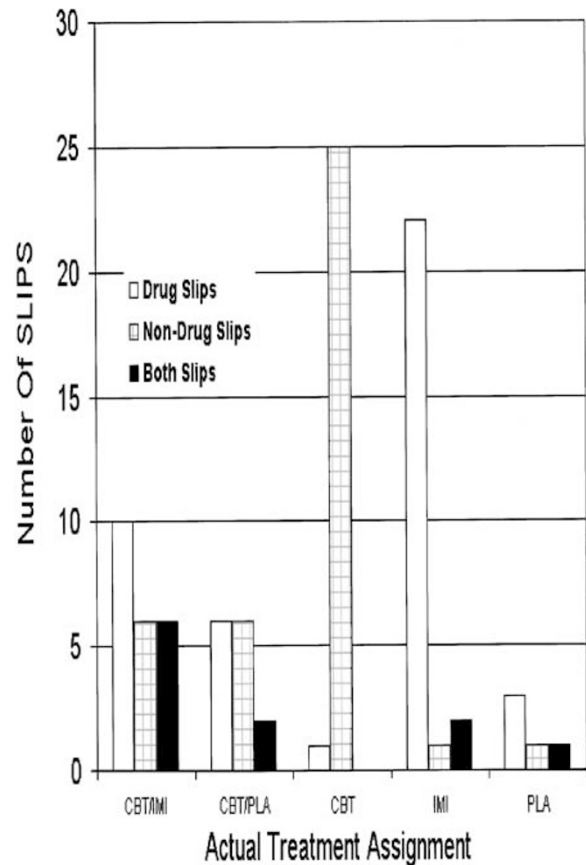

Figure 2 Type and number of SLIPS made by patients in different treatment conditions.

Table 2 shows the distribution of IE correct guesses categorized by patients' response to treatment. Data are presented only for the PA assessment since there were no differences observed across evaluation points.

Table 3 shows the MIXOR regression analysis applied using four dummy variables to represent the different treatment assignments and actual treatment assignments using dummy variables (IMI, CBT, PLA, CBT/IMI). We ordered the coefficients to determine the relative difficulty of determining the correct treatment assignment. These results showed correct guessing according to actual treatment assignment in the following order arranged from highest to lowest rates of correct guessing (CBT, IMI, CBT/IMI, PLA, CBT/PLA)

Figure 4 shows the accuracy of IE guesses about assignments to CBT conditions and active medication conditions at PA allowing direct comparison with similar data presented by Carroll et al (1994). Applying the mixed-effect model to subgroupings, we found that rates of correct guessing by IEs were not significantly associated with assignment to behavior therapy (CBT/IMI, CBT/PLA or CBT; $p = 0.16$) or active medication (CBT/IMI, IMI; $p = 0.08$).

DISCUSSION

Several procedures were implemented to provide objective and independent evaluations of the relative effectiveness of

Figure 1 Percent correct classifications of treatment groups by IEs at different treatment sites for PA, PM, and FF assessments.
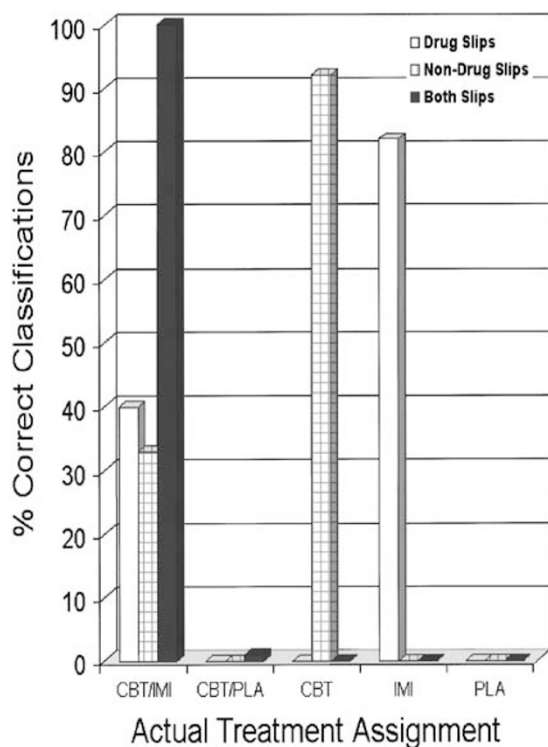
Figure 3 Percent correct classifications of treatment conditions by IEs recording different types of SLIPS by patients in different treatment conditions.

Table 3 Mixed-Effects Regression Model of IE Correct Guesses about Treatment Assignment on Assessment Time, Treatment and Patient Response to Treatment

	Fixed effects			
Variable	**Estimate**	**Standard error**	**Z**	***p*-value**
Intercept	−3.35	0.67	−4.96	0.000
Postacute	−0.51	0.40	−1.28	0.20
Postmaintenance	−0.20	0.39	−0.51	0.61
SLIPS	1.06	0.43	2.48	0.01
IMI	0.32	0.60	5.35	0.000
PLA	2.17	0.68	3.19	0.001
CBT/IMI	2.60	0.60	4.35	0.00001
CBT	3.96	0.66	5.99	0.000
Response to treatment	0.58	0.38	1.53	0.13

medical and psychosocial treatments for panic disorder with agoraphobia. In general, these procedures were not effective in keeping IEs from accurately guessing the treatment group assignment of patients. This was a robust finding unaffected by individual IEs, assessment point (PA, PM, or FF), profession of the evaluator, or patients' response to treatment.

A significant association was found between treatment site and correct guessing by IEs. IEs at sites 2, 3, and 4 showed a significantly greater association between correct guesses and actual treatment assignment than IEs at Site 1. It is of interest that the rank order of correct guessing across sites is related to the IEs' level of involvement in the setting.

Site 1 IEs were consultants who were on campus only to conduct the independent evaluations. At Site 3, the IEs were part of the project staff, but worked in a different building on the same campus. At the other two sites, IEs were full-time staff members who had ample opportunity to witness patients entering and exiting from therapists' offices and therefore make a connection between the therapist and the treatment that the patient was receiving.

The IEs were not involved in treatment or research other than the assessments. Prior to all post-treatment evaluations, research staff, therapists, and IEs asked patients not to provide information about treatment to the IEs during these interviews. In spite of these precautions, patients frequently slipped and revealed this type of information. Not surprisingly, this significantly improved IE's ability to guess correctly. Analysis of the data on SLIPS clearly shows that the information provided by SLIPS was associated with the treatment condition and enhanced the likelihood of correct IE guesses.

Although the statistical analysis found no significant association between patients' response to treatment and IE correct guessing, the data show an apparent bias on the part of the IEs that treatments worked. IEs never correctly

Table 2 IE Classifications of Treatment at the PA Evaluation as a Function of Response to Treatment Status

	IE classifications of treatment groups					
	Responders			**Nonresponders**		
Treatment group	**Correct**	**Incorrect**	**% Correct**	**Correct**	**Incorrect**	**% Correct**
CBT/IMI	18	15	54.55	0	5	0.00
CBT/PLA	2	29	6.45	0	5	0.00
CBT	23	6	79.31	6	5	54.55
IMI	19	12	61.29	6	6	50.00
PLA	0	7	0.00	4	2	66.67
TOTAL	62	69	47.33	16	23	41.03

IE Correct Guesses on Assignment To Behavior Therapy (CBT/IMI,CBT/PLA or CBT)Overall: 127/170(n) = 74.7% Correct

No Medication: 65/89 = 73.0% (CBT/PLA CBT, PLA)

No CBT: 9/13 = 69.2% (PLA)

CBT: 61/76 = 80% (CBT/PLA, CBT)

Medication: 57/81 = 70.4% (CBT/IMI, IMI)

No CBT: 31/43 = 72.1% (IMI)

CBT: 26/38 = 68.4% (CBT/IMI)

IE Correct Guesses on Assignment To Active Medication (CBT/IMI, IMI) Overall 111/170 =65.3% Correct

No CBT: 37/56 = 66.1% (IMI, PLA)

No Medication 6/13 = 46.2% (PLA)

Medication: 31/43 = 72.1% (IMI)

CBT: 74/114 = 64.9% (CBT/IMI, CBT/PLA, CBT)

No Medication: 47/76 =61.8% (CBT/PLA, CBT)

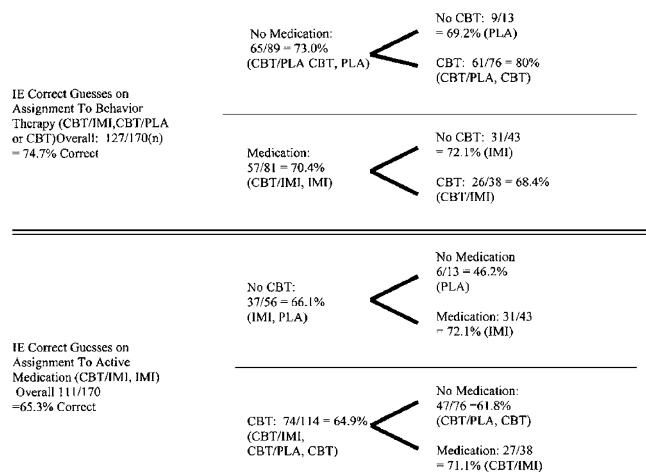Medication: 27/38 = 71.1% (CBT/IMI)

Figure 4 Accuracy of IE guesses about patient treatment conditions for subcategories of treatment assignment.

classified patients with positive responses to treatment into the PLA group. Furthermore, IEs did not correctly classify 10 nonresponders into CBT or CBT/PLA groups. The fact that the IEKNO was completed immediately after IEs completed the CGI Severity and Improvement Scales certainly allows for the influence of outcome ratings on IE guesses. It is possible that this association would appear even stronger in a study simply comparing an active treatment to a placebo or in a study with more placebo subjects. Assessing base rates of this pre-existing bias may be a more accurate way to control for 'blindness' of the IE—especially since similar accuracy rates have been reported in several independent studies. The tendency to choose CBT over CBT/PLA was a response bias observed across IEs. The relatively low number and confusing pattern of SLIPS recorded for those patients may have contributed to this bias.

The results of this study support the findings by Carroll et al (1994) in that IE guesses were correct at greater than chance rates for both psychotherapy and pharmacotherapy. In fact, the accuracy rates for the two studies are remarkably similar. While the present study found an interaction between treatment conditions and correct guessing, we did not find the differential accuracy of guessing for patients assigned to medication groups vs those assigned to nonmedication treatments that Carroll et al (1994) reported. Our overall findings were also quite similar to Basoglu et al (1997) including the fact that both studies did not find a relationship between treatment outcome and correct guessing by IEs.

There are several limitations to the current study. First, results are based on approximately two-thirds of the patients who participated in post-treatment evaluations and the selection of that sample was not random. The IEKNO was initiated after the first year of the study in response to frequent IE reports of potentially blind-breaking information. So, it is not known if this sample is an accurate representation of the population. Second, we assumed that because there were five treatment conditions and the IEs were not told of the unequal distribution of patients across conditions that 20% of the IE guesses would occur in each category by chance. It is not possible to know

from this study whether the assumption of equal distribution is valid. Also, it is possible (as with other aspects of the blinding procedures) that some IEs did know about the actual distribution of patients into groups. Third, the categorization of information into types of SLIPS/NON-SLIPS by raters after the fact may not accurately reflect the type or source of information provided by IEs in response to an open-ended question. In future research, it would be preferable to list specific types of information known to contribute to enhancing correct guessing and have IEs select the type of information from that list. The IEKNO could also be enhanced by asking IEs whether their selection of treatment condition for a patient was based on response to treatment, additional information, or both. Fourth, this evaluation of blindness to treatment would have been enhanced by reporting the effectiveness of the double-blind procedures for patients, pharmacotherapists, and IEs in the medication conditions (Bakker et al, 1999; Basoglu et al, 1997). Attempts were made to collect this information, but due to a number of factors we were not able to gather data sufficient for valid conclusions. Fifth, it is possible that use of the CGI-S and CGI-I to define responder status could have been more prone to rater bias than a measure such as the PDSS that has more specific behavioral referents. Indeed, Carroll et al (1994) reported that subjective outcome ratings were more related to accuracy of IE guesses than were more objective measures. Although each evaluation included a PDSS, we chose to report treatment response here using the primary definition for responder status used in the original study (Barlow et al, 2000).

CONCLUSIONS AND RECOMMENDATIONS

This study adds to a growing literature indicating that double-blind procedures are often not completely effective in protecting IEs from accurate guessing about treatment assignments of study participants. Thus, it is possible that some of the positive patient response to medication or relative lack of response to placebos could be attributed to expectancies. The results of the original study (Barlow et al, 2000) need to be evaluated in light of this finding. The fact that IEs guessed treatment conditions at a rate significantly above chance raises the concern that an assumption about the patient treatment group would influence IE ratings about the outcome of treatment.

Several aspects of the Barlow et al (2000) study were designed to provide protection against the impact of a possible systematic bias in responding by the IEs. On a large scale, the design and close monitoring of the study by principle investigators who represented alignment to both medical and psychological interventions, the use of multiple sites, and the use of IEs who represented both medical and nonmedical professions reduced the likelihood of a systematic bias. A second level of procedures designed to control for systematic bias occurred at the IE level and consisted of training to criteria on the primary outcome measures as well as periodic independent monitoring of the evaluation implementation and scoring. A third level of protection was provided by the relatively objective nature of the ratings upon which effectiveness of treatment were based.

The most enlightening finding of this study comes from the discovery that, in spite of the many procedures to protect the blindness of IEs, both patients and project staff provided clues to treatment assignment. If study staff and patients are more effectively alerted to this problem, it seems likely that procedures could be devised to reduce the occurrence of slips. Some of these clues could be eliminated or reduced by reducing contact between IEs and treatment/research staff. When possible, IEs should conduct their interviews in locations that do not allow them to see patients attending treatment. IEs appear less able to guess treatment assignment correctly when the assessment interviews are their only association with the project.

Research staff can help uphold the blind by not revealing to IEs therapist names, treatment conditions, or responder status, and access to paperwork or codes identifying treatment conditions. IEs should not attend meetings where treatment issues and patients are discussed. Research staff, therapists, and IEs can play an important role by giving more specific instructions to patients about what not to reveal to the IEs during post-treatment interviews. Patients should be asked specifically not to reveal their therapist/physician name or characteristics, effects or side effects of medications, coping strategies, relaxation techniques, or attempts to challenge their fears. This detail is easily overlooked, but plays a significant role in breaking of the blind. One well-meaning patient in response to being asked by an IE not to reveal any clues about her treatment exclaimed 'Of course, Dr. (CBT therapist's name) just told me the same thing.'

When an IE guesses correctly due to the assumption that a responder is in an active treatment condition and a nonresponder is in a placebo condition, this situation does not necessarily lead to bias. However, when an IE guesses the treatment assignment due to a 'SLIP', bias enters the picture if that correct guess influences the ratings of treatment efficacy. Thus, it is important to utilize information about procedures that can increase independence of evaluators and reduce the opportunities for bias to intrude on determinations of treatment efficacy. Additional research is needed to develop methods for assessing bias in estimating treatment efficacy when a blind has been compromised.

ACKNOWLEDGEMENTS

REFERENCES

Bakker A, Spinhoven P, Anton J, Dick M, van Dyck R (1999). Double-blindness procedure did not mask giving of medication in panic disorder. *J Aff Dis* **54**: 189–192.

Barlow D, Gorman J, Shear K, Woods S (2000). Cognitive-behavioral therapy, imipramine, or their combination for panic disorder. *JAMA* **283**: 2529–2536.

Basoglu M, Marks I, Livanou M, Swinson R (1997). Double-blindness procedures, rater blindness, and ratings outcome: observations from a controlled trial. *Arch Gen Psychiatry* **54**: 744–748.

Beatty W (1972). How blind is blind? A simple procedure for estimating observer naivete. *Psychol Bull* **78**: 70–71.

Carroll K, Rounsaville B, Nich C (1994). Blind man's bluff: effectiveness and significance of psychotherapy and pharmacotherapy. blinding procedures in a clinical trial. *J Consult Clin Psychol* **62**: 276–280.

Chambless DL, Hollon SD (1998). Defining empirically supported therapies. *J Consult Clin Psychol* **66**: 7–18.

DiNardo PA, Barlow DH (1988). *Anxiety Disorders Interview Schedule-Revised (ADIS-R)*. Graywind Publications Inc/The Psychological Corporation: San Antonio, TX.

Double D (1995). Unblinding in trials of the withdrawal of anticholinergic agents in patients maintained on neuroleptics. *J Nerv Ment Dis* **183**: 599–602.

Fisher S, Greenberg R (1993). How sound is the double-blind design for evaluating psychotropic drugs? *J Nerv Ment Dis* **181**: 345–350.

Gibbons RD (2000). Mixed-effects models for mental health services research. *Health Services Outcomes Res Methodol* **1**: 91–129.

Guy W (1976). *ECDEU Assessment Manual for Psychopharmacology, Revised*. US Government Printing Office: DHEW Publication ADM: Washington DC, 76–338.

Hedeker D, Gibbons RD (1996). MIXOR: a computer program for mixed-effects ordinal regression analysis. *Comput Meth Programs Biomed* **49**: 157–176.

Kendall PC (1998). Empirically supported psychological therapies. *J Consult Clin Psychol* **66**: 3–6.

Klein DF (1994). The utility of guidelines and algorithms for practice. *Psychiatr Ann* 362–367.

Marks I, Gray S, Cohen D, Hill R, Mawson D, Ramm E, Stern R (1983). Imipramine and brief therapist-aided exposure in agoraphobics having self-exposure homework. *Arch Gen Psychiatry* **40**: 153–162.

Schneier FR, Goetz D, Campeas R, Fallon B, Marshall R, Liebowitz MR (1998). Placebo-controlled trial of moclobemide in social phobia. *Br J Psychiatry* **172**: 70–77.

Seligman MEP (1995). The effectiveness of psychotherapy: the consumer reports survey. *Am Psychol* **50**: 965–974.

Shear MK, Brown TA, Barlow DH, Money R, Sholomshas DE, Woods SW, Gorman JM, Papp LA (1997). Multicenter collaborative panic disorder severity scale. *Am J Psychiatry* **154**: 1571–1575.

Spitzer RL, Williams JBW, Gibbon M, First MB (1988). *Structured Clinical Interview for DSM-IIIR: Patient Version (SCID-P)*. New York Psychiatric Institute: New York.

Woods SW, Sholomskas DE, Shear MK, Gorman JM, Barlow DH, Goddard AW *et al.* (1998). Efficient allocation of patients to treatment cells in clinical trials with more than two treatment conditions. *Am J Psychiatry* **155**: 1446–1448.