



100
815
THS

2000 1553



This is to certify that the

thesis entitled

Client Satisfaction and
Meaningful Change in Psychotherapy

presented by

George Yunus Ankuta

has been accepted towards fulfillment
of the requirements for

M. A. Psychology
degree in

A handwritten signature in cursive script, appearing to read "H. M. Wells", written over a horizontal line.

Major professor

Date 5/19/88



RETURNING MATERIALS:

Place in book drop to
remove this checkout from
your record. FINES will
be charged if book is
returned after the date
stamped below.

--	--	--

CLIENT SATISFACTION AND
MEANINGFUL CHANGE IN PSYCHOTHERAPY

By

George Yunus Ankuta

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Psychology

1988

ABSTRACT

CLIENT SATISFACTION AND MEANINGFUL CHANGE IN PSYCHOTHERAPY

By

George Yunus Ankuta

The purpose of this study was to evaluate the use of "clinical significance" in psychotherapy data analysis. Clinical significance was defined, in part, whether the client's scores on a symptom measure move from the dysfunctional to the functional range at the termination of treatment. Seventy-five adult psychotherapy clients were divided into three groups based on degree of psychological disturbance. The hypotheses investigated were: 1) a group of psychotherapy clients showing clinically significant symptom changes will report greater satisfaction and benefit from psychotherapy than a group of clients whose changes fell solely within the statistically significant improvement range, and 2) psychotherapy clients showing statistically significant improvement will report greater satisfaction and benefit from psychotherapy than a group of clients who did not improve statistically or clinically. The first hypothesis was supported. It is suggested that using clinical significance in psychotherapy data analysis is a

George Yunus Ankuta

way of bridging the researcher-practitioner gap by
providing a measure of meaningful change which appears to
have social validity.

To my parents

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my thesis committee, Dr. Norman Abeles, Dr. Bertram Karon, and Dr. Raymond Frankmann. Dr. Abeles' student-centered approach created an environment in which I enjoyed developing my ideas. His personal support, acceptance, and respect helped make this research possible. His enthusiasm and expertise in psychotherapy research have been inspiring.

In addition I would like to express my special appreciation to Lisa Cowden whose love and warmth has added so much to my life.

TABLE OF CONTENTS

	<u>Page</u>
LIST OF TABLES	v
LIST OF FIGURES	iv
Introduction	1
Statistical and Clinical Significance Defined	1
Issues with the Use of Statistical Significance	4
When has Meaningful Change Occurred in Psychotherapy 8	
Research Reasoning Should Serve the Needs of The	
Clinical Field it Intends to Study.	15
Research Reasoning Should Match the Reasoning of	
the Clinical Field it Intends to Study	20
Hypotheses.	27
Method	29
Subjects	29
Materials	30
Symptom Check List 90 Revised	30
The Strupp Post Therapy Client Questionnaire	34
Procedure	36
The Groups.	38
Operational Definitions of Group Criteria	38
Results.	43
Discussion.	49
References.	57

LIST OF TABLES

<u>Table</u>	<u>Page</u>
1 Symptom Check List 90 Revised Data Used in Determining Clinical Significance Outcome Criteria.	40
2 The Relationship of Statistical Significance and Clinical Significance of Level of Symptom Change to Client Satisfaction.	46
3 Pairwise Contrasts of Groups by Client Satisfaction	47
4 Group Differences on Pre-therapy Symptom Level as Measured by SCL-90-R Scales	48

LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
1	Hypothetical Data From an Imaginary Measure Used to Assess Change in a Psychotherapy Outcome Study.	3

Introduction

Statistical and Clinical Significance Defined

Statistical significance refers to the evaluation of parameters of distributions using statistical hypothesis testing (Hays, 1981, ch. 7). Clinical significance refers to the effect of a treatment procedure on a single subject (Hugdahl & Ost, 1980).

"Clinically significant change has been defined as a large proportion of clients improving (Hugdahl & Ost, 1981), a change which is large in magnitude (Barlow, 1981), an improvement in the client's everyday functioning (Kazdin & Wilson, 1978), a change which is recognizable to peers and significant others (Kazdin, 1977; Wolf, 1978), an elimination of the presenting problem (Kazdin & Wilson, 1978), and the attainment of a level of functioning which is no longer distinguishable from the client's nondeviant peers (Kazdin & Wilson, 1978; Kendall & Norton-Ford, 1982)." (Jacobson et al., 1984, p. 338).

Jacobson, Follette, & Revenstorf (1984) propose a two-condition evaluation to judge the criterion of clinical significance. The first condition is a measure of meaningful change which checks whether the client has moved from the dysfunctional to the functional range. The second condition is whether or not the improvement is statistically reliable.

The condition of meaningful change is evaluated by asking if the level of functioning posttreatment suggests that the subject is statistically more likely to be in the functional than in the dysfunctional population. In other words, is the posttreatment score statistically more likely to be drawn from the functional than the dysfunctional distribution? There is a point we will call "c" where the probabilities of belonging to the functional and dysfunctional populations are equal. If the pretreatment and post treatment distributions are symmetric and of the same shape or are mirror images through the ordinate at "c", then "c" can be determined mathematically by the following equation:

$$c = \frac{S_0 \bar{X}_1 + S_1 \bar{X}_0}{S_0 + S_1} \text{ (see note 1 and figure 1)}$$

If X_{post} is greater than "c" the client is more likely to be in the functional population. If X_{post} is less than "c" the client is more likely to be in the dysfunctional population.

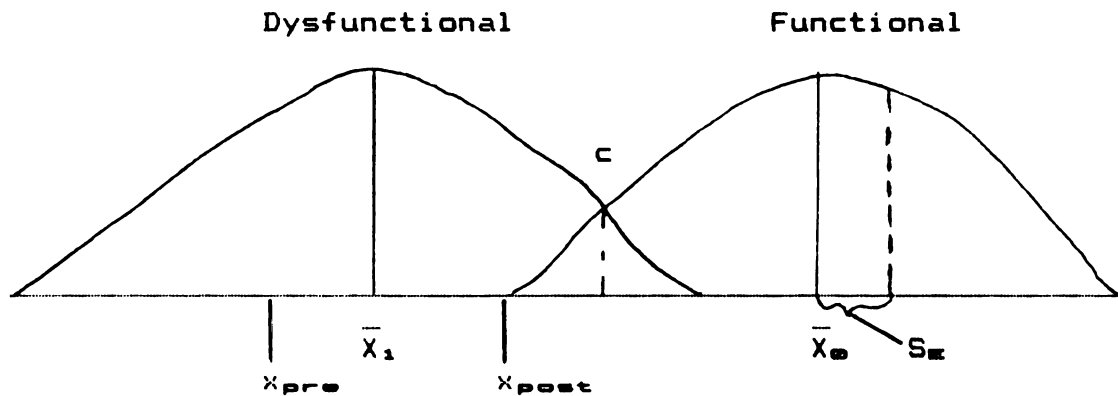
The condition of reliable change is evaluated by asking if the Reliable Change Index (RC) (Jacobson et al. 1984), defined by the following formula, is greater than 1.96 (see note 1):

$$RC = (X_{\text{post}} - X_{\text{pre}}) / S_E$$

$$S_E = S_1 \sqrt{1 - r_{xx'}}$$

Figure 1

Hypothetical Data From an Imaginary Measure Used to Assess
Change in a Psychotherapy Outcome Study. (Jacobson et al.
1984)



Note 1:

- \bar{X}_1 = mean of both pretreatment experimental and pretreatment control group
- \bar{X}_0 = mean of the well functioning population
- x_{pre} = pretreatment score of a hypothetical subject
- x_{post} = posttreatment score of a hypothetical subject
- S_1 = standard deviation of the pretreatment control group, and pretreatment experimental group
- S_0 = standard deviation of the well functioning population
- $r_{xx'}$ = test-retest reliability of this measure for a dysfunctional sample
- S_e = standard error of measurement of this measure

Issues with the Use of Statistical Significance

Though statistical significance tests of parameters of distributions are the prominent method of treatment evaluation, the way they are sometimes used is subject to shortcomings. Several articles recommend supplementing statistical significance tests with tests of clinical significance and suggest measures of clinical significance (Hugdahl & Ost, 1981; Jacobson et al. 1984; Lick 1973).

There are weaknesses in the way statistical significance tests are used. There is really no good reason to expect the null hypothesis to be true in any population. Examination of any set of statistics on a total population will quickly confirm the rarity of the null hypothesis in nature (Bakan, 1966). Rejecting something that is not likely to be true should not be the end goal of data analysis, although the null hypothesis still remains a potential explanation of any finding and it needs to be disposed of.

Misconceptions about statistical significance testing lead to its inappropriately being used alone for presenting psychotherapy outcome research data. The "odds against chance fantasy" is the misinterpretation of the p

value as the probability that the research results were due to chance or caused by chance (Carver, 1978). The "p value" of a statistic is not a probability. After all, when a statistic is computed one has a number, not a random variable. The p value is calculated by assuming that some specific chance process did produce the mean difference, and the p value is used to decide whether to accept or reject that assumption (Carver, 1978).

Another misconception is the "replication or reliability fantasy", which is the misinterpretation of statistical significance as the probability of obtaining the same result whenever a given experiment is repeated. Nothing in the logic of statistics allows this inference (Carver, 1978). Significance obtained ($p < .05$) does not mean that if the experiment were repeated 100 times the same difference would occur 95 out of 100 times (Hugdahl & Ost, 1981). It does mean that if the null hypothesis is true, the probability of outcomes in the alpha level rejection region is alpha.

Yet another misconception is the "valid research hypothesis fantasy" which involves concluding the research hypothesis is true as a result of statistical significance tests of parameters of distributions (Carver, 1978). Scientific hypotheses are different from statistical hypotheses and require more than statistical significance

tests in one experiment to support them (Bolls, 1962; Winch, 1969). The scientist uses statistical hypotheses and tests to investigate a scientific hypotheses about nature. A statistical test in an experiment, significant or not, is merely one piece of evidence in the scientist's attempt to determine what is true about the natural world and establish support for his view. The fit of the statistical model, the plausibility of alternative hypotheses, and all available data must be considered before the scientific hypothesis can be evaluated. Statistics don't know where the numbers come from, but it is up to the scientist to know.

There are limitations on the way statistical significance tests should be used for evaluating psychotherapy outcome data: 1) Statistical significance testing can draw attention from the practical question of the applied importance of behavior change (Kazdin, 1980). 2) Significance testing of parameters of distributions does not usually convey information about single subjects within the sample tested. Significance obtained ($p < .05$) does not mean that if any randomly selected subject from group A receiving treatment A is compared with any randomly selected subject from group B receiving treatment B the difference will exist 95% of the time in the predicted direction. What is the probability

that treatment A is better than treatment B for a particular client? This cannot be determined by statistical significance testing (Hersen & Barlow, 1976; Hugdahl & Ost, 1981).

Another misuse of statistical significance observed by Kazdin (1978) is that the experimenter may use statistical significance in a way that obscures important information. Suppose two treatment groups are compared. In group A, 2 of 10 change a large amount and in group B, 8 of ten change a small amount. Statistical significance tests of parameters of distributions will not differentiate these data unless they are specifically designed to do so.

Many of the issues with the use of statistical significance testing mentioned above arise in psychotherapy outcome research as a result of practitioner's need to know about the variability in improvement data so that they can select the best treatments for their clients. When variability is not reported readers may use inappropriate means of estimating variability. Therefore some aspect of the data analysis must provide information on that variability. Clinical significance, which is evaluated on a case by case basis, gives information on the variability of psychotherapy outcome data and is suggested for that

purpose. Clinical significance is an apt augmentation to statistical significance tests which provides additional information that practitioners need.

When has Meaningful Change Occurred in Psychotherapy?

Meaningful change has occurred when change enhances the client's everyday functioning. Reduction of hand washing behavior in a client with obsessive compulsive symptoms from 100 to 70 occurrences per hour is not overwhelmingly meaningful from a clinical perspective since it does not significantly enhance everyday functioning. If after treatment the individual's behavior falls within normal levels of nonproblem peers, meaningful change has occurred. The extent to which treatment restores adequate levels of functioning needs to be assessed directly (Kazdin & Wilson, 1978).

A unique concern of clinical research is in effecting changes in the client that are clinically significant or actually make a difference in the client's life. Clinically meaningful changes should be dramatic and obvious from the data so that there is no need to refer to statistical tests (Kazdin, 1977). How do we judge if a change is clinically important? One way to evaluate clinical importance of a change is to consider its social validity.

Judgments about social validity are made in reference to the following: "1) The social significance of the goals. Are the specific behavioral goals really what society wants? 2) The social appropriateness of the procedures. Do the ends justify the means? That is, do the participants, care givers and other consumers consider the treatment procedures acceptable? 3) The social importance of the effects. Are consumers satisfied with the results? All the results, including any unpredicted ones?" (Wolf, 1978).

How would social validity be evaluated? Kazdin (1977, 1980) discusses two ways that have been used: The social comparison method and the subjective evaluation method. In the social comparison method the behavior of the client is compared to the behavior of "nondeviant" peers. The question asked is whether the client's behavior after treatment is distinguishable from the behavior of his peers. The social comparison method was used in the social validation and training of conversation skills (Minkin et al. 1976). Junior High School girls were trained in three aspects of effective conversation. When judged against their nontrained peers they were rated as superior in conversation skill.

In the subjective evaluation method the client's behavior is evaluated by individuals who are likely to

have contact with that client to determine whether the change made during treatment is significant. This method has been used by Patterson (1974) in a study of interventions for boys' conduct disorder problems. Direct observations were made in the boys' homes and classrooms before, during, and after the intervention. Daily reports on the boys' problem behavior were obtained from their parents. Changes in the problem behavior were accompanied by consistent changes in the parents' perceptions.

The crucial factor in the social comparison method is to identify the client's peers. The peers are those individuals similar to the client in subject and demographic variables, but different in performance of the target behaviors. There are two ways the peer group can be used: 1) All individuals in a situation (i. e. classroom) can be used to determine whose behavior is extreme 2) The level of behavior of the peers who did not warrant treatment can serve as the criterion by which to assess the success of treatment. If treatment has been successful the client's performance should fall within the normative level of his peers (Kazdin, 1977). Normative judgements should be incorporated directly into treatment evaluation.

There are several considerations in using normative

data. Occasionally normative standards are inadequate. Sometimes it is the norm that must be changed. For example, classroom performance in an entire school may be too low, or waste disposal in an entire industry may be unsatisfactory. Another consideration is that identifying the normative group can be difficult. In evaluating retarded patients should the normal group be society normals or untreated retarded people. In a prison situation should the norm group be nontreated prisoners or nonprisoners (Kazdin, 1977)?

In addition to determining whether society would be satisfied with the individual's change, meaningfulness of change in psychotherapy can be evaluated by considering whether the individual, and the mental health practitioner, are satisfied with the changes that the individual has made in therapy. Taken together, these three perspectives on mental health, of society, the client, and the practitioner, constitute Strupp's (1977) Tripartite model of mental health and therapeutic outcome. The model highlights the values brought to bear by the three "interested parties." Mental health and favorable therapeutic outcome can only be achieved when all three "interested parties" are satisfied. What more could the client, the practitioner and society want then the client's return to normal functioning? Returning to

the range of normal functioning is an intuitively appealing and nonambiguous measure of therapeutic outcome especially if one is not overly obsessive about defining the meaning of normality.

Another way to evaluate the meaningfulness of change in therapy is to consider the magnitude of the change. Garfield (1981) suggests that we must look beyond standard significance tests to the extent of change in therapy. He suggests that studies in which the posttreatment outcome measure did not exceed the mid range of the scale are not meaningful, though statistical significance may be achieved, because there is still so much room for improvement. He claims that only large change is clinically meaningful.

Cronbach and Furby (1970) discuss the use of change scores to evaluate the outcome of treatment. Persons may differ on the posttreatment measure more than predicted from the pretreatment score using regression. These positive deviations are subject to many competing explanations other than that these individuals benefited particularly well from treatment. They may have started with some valuable attribute that the pretreatment measure did not encompass. The pretreatment score may have been underestimated. The posttreatment scores may have been overestimated. Their posttreatment success may be an

accidental effect arising from some tactic casually adopted during treatment.

Cronbach suggests that most often it is best to use the pretreatment and posttreatment scores as two variables separately in the analysis to allow for more complex relationships. A very disturbed patient for example may improve because of the large distance needed to achieve adequate functioning. Level of pathology rather than magnitude of change conveys the information. A neurotic patient may not show as much improvement because as one becomes more functional change may be less noticeable. Much of the important information is in pretreatment scores. Cronbach suggests that investigators who ask questions about gain scores would ordinarily be better served by framing their questions in other ways.

Pretreatment level of pathology may be a factor in improvement in therapy, and alternative explanations for change are always possible, however the effects of pre and post treatment symptom level can be evaluated independently of change. The effect of the level of change on client satisfaction is important in itself. Obviously a large change cannot be achieved in an individual with a low level of pathology. A criterion for the evaluation of treatment that identifies those most satisfied with therapy is an important and useful criterion to establish.

The measure of clinical significance suggested by Jacobson et al. (1984) satisfies all the notions of meaningful change discussed above. To meet Jacobson et al.'s criterion of clinical significance the participant must be statistically more likely to be in the normal than the abnormal distribution of scores, and the change must be large enough to be statistically reliable (more than two standard deviations). Requiring that the participants be more likely to be in the normal than the abnormal group guarantees that they will satisfy Kazdin's condition of social validity. The social comparison method would demonstrate that the participant is like his or her nondeviant peers because now he/she is likely to be nondeviant. The subjective evaluation method would yield favorable ratings from those who are likely to have contact with the participant because now he/she is likely to be more functional and "normal". If the client returns to the normal range of functioning, client, society, and the mental health practitioner will all be satisfied with the outcome (Strupp, 1977). Jacobson et al.'s criterion of reliable change guarantees that Garfield's (1981) criterion, that meaningful change is change of great magnitude will be satisfied.

Research Reasoning Should Serve the Needs of the Clinical Field it Intends to Study.

Recent trends in clinical psychology training have tended to institutionalize the research-practitioner split. Professional programs that stress the development of clinical skill exclusively and leave the research to the Ph. D.s are becoming more prominent. Previously it was observed that clinicians are unlikely to engage in research of any kind. The modal number of publications of clinical psychologists is zero (Kelly, Goldberg et. al. 1978). More seriously, many clinicians are not influenced by clinical research findings (Barlow, 1981).

There are technical problems that complicate clinical research and interfere with relevance. It is hard to collect clients in large numbers that is homogeneous for a particular behavior disorder. This makes it difficult to test hypotheses about groups of people with particular disorders. There are ethical objections to withholding treatment from clients for research purposes: 1) Control groups are unethical because some persons are deprived of the treatment that they need. 2) It is impossible to insure that persons in the control group will not seek help from other

professionals and friends etc. (Smith et al., 1980). This makes establishing a control group difficult. "Psychotherapy is complex and not standardized; no two clients are treated the same way by even the same psychotherapist; so psychotherapy cannot be labeled method A or method B and studied experimentally." (Smith et al., 1980, p. 28). It is difficult to answer questions about which treatment for which individual with the group comparison research strategy that is currently popular (Barlow, 1981).

Procedural and philosophical differences exist between researchers and practitioners that make functioning in both modes difficult. Practitioners tailor the length, intensity and method of any investigation to the individual and his problem. Researchers tend to continue treatment guided primarily by the experimental condition. The lack of emphasis on the individual diminishes the importance and relevance of clinical research for the practitioner.

Further evidence for the researcher-practitioner split is the lack of clinical relevance of clinical research. However clinical research is defined, one ingredient of the definition has to be clinical relevance. Can the content of a research article be in some way used to help the patient.? Maletzky (1981) did a

study of all the articles published in ten journals selected for their high circulation rate and their expressed goal of providing useful information to practitioners. All issues of the journals from January 1978 to June 1980 were reviewed. It was found that only 25.1% of the psychology journals and 17.3% of the psychiatry journals contained any immediately useful information bits for the clinical situation. A "bit" was arbitrarily defined as a "practical" unit of information that could be used to treat patients directly. Only 4.3 "bits" of clinically useful information were contained in the average psychology journal, and only 1.3 bits per issue were contained in the average psychiatry journal. It was estimated that approximately 48.8 minutes of reading time were consumed for each "bit" of clinical information gleaned from a psychology journal and over three hours were consumed to get each "bit" of clinical information from a psychiatry journal. How well is clinical research serving the needs of the clinician?

Strupp (1981) writes about the crisis of confidence facing psychotherapy today. There is a public demand for better scientific evidence on efficacy and safety of psychotherapy. Society is demanding accountability of mental health practitioners. The profession must be able to articulate to the public in acceptable and

understandable terms that we have the means to help them.

Strupp (1981) comments that in our quest for knowledge about human interaction each therapeutic dyad constitutes an experiment. Young therapists can learn only from the study of individual cases. If we carry out group comparisons without sustained attention to the process in individual dyads we deprive ourselves of the most important opportunity that systematic research has to offer.

We can better learn about psychotherapy and persuade the public if we use clinical significance to evaluate therapy outcome. Clinical significance has an intuitive interpretation and can be explained to the public. In addition, it provides a means for each therapist to evaluate himself and be held accountable for his or her therapy outcomes.

Cronbach (1975) comments that the historic separation of experimental psychology from the study of individual differences impeded psychology research. Some 30 years ago, research in psychology became dedicated to the quest for nomothetic theory. Model building and hypothesis testing became the central concern. Research problems were chosen to fit that mode.

Cronbach (1975) suggested an alternate mode of inquiry: the mode of intensive local observation.

"...An observer collecting data in one particular situation is in a position to appraise a practice or proposition in that setting, observing effects in context. In trying to describe and account for what has happened, he will give attention to whatever variables were controlled, but will give equally careful attention to uncontrolled conditions, to personal characteristics, and to events that occurred during treatment and measurement. As he goes from situation to situation, his first task is to describe and interpret the effect anew in each locale, perhaps taking into account factors unique to that locale of series of events (cf. Geertz, 1973, chap. 1, on "thick description"). As results accumulate, a person who seeks understanding will do his best to trace how the uncontrolled factors could have caused local departures from the modal effect. That is, generalization comes late, and the exception is taken as seriously as the rule." (Cronbach, 1975, p. 124-125).

Barlow (1981) notes that intense local observation could be a way of closing the research-practitioner gap. It would provide clinicians with more clinically relevant information such as what type of treatments work for which type of individual. Clinicians could be more actively involved in research. Clinicians could collect data on hundreds of thousands of cases over several years. The information could be fed into large clinical research centers (Argas, Kazdin & Wilson, 1979). This would make clinicians and researchers more interdependent.

Clinicians prefer studies that tell about the clinical significance of the findings. In a survey (Sargent, 1983) of 530 members of The American Psychological Association Division 37 (Child, Youth, and Family Services) in which respondents were asked to rate

versions of a psychotherapy research study, experimental versions of the design received higher ratings than quasi-experimental versions and nonexperimental versions. The versions that reported the finding's clinical significance received a higher methodology rating than versions that omitted this information. Practitioners would prefer research that meets their needs. Practitioner's prefer research that considers clinical significance.

Research Reasoning Should Match the Reasoning of the Clinical Field it Intends to Study

When psychology and physiology became sciences, the initial experiments were performed on individual organisms and the results of these initial investigations remain relevant to the scientific world today. Broca examined a man who was unable to speak. When performing an autopsy after death Broca discovered a lesion in the third frontal convolution of the cerebral cortex. He determined it was the speech center of the brain and it is now named after him. Pavlov's basic findings were gleaned from single organisms and strengthened by replications in other organisms (Hersen & Barlow, 1976).

The study of individual differences and the

statistical approach to psychology became prominent during the first half of the twentieth century. With a push from the functional school of Psychology, and a developing interest in measurement and testing of intelligence, the foundation for comparing groups of individuals was laid. Galton and Pearson expanded the study of individual differences at the turn of the century and developed many of the descriptive and inferential statistics still in use today (Hersen & Barlow, 1976).

"It may seem ironic at first glance that a concern with individual differences lead to an emphasis on groups and averages, but differences among individuals, or inter-subject variability, and the distribution of these differences necessitate a comparison among individuals and a concern for a description of a group or population as a whole. In this context observations from a single organism are irrelevant." (Hersen & Barlow, 1976, p. 6).

There are many advantages in clinical research to single case experimental designs. For example, "attempts to apply an ill-defined and global treatment such as psychotherapy to a heterogeneous group of clients classified under a single diagnostic category such as neurotics are incapable of assessing the more basic question on the effectiveness of a specific treatment for a specific individual." (Hersen & Barlow, 1976, p.13). Single case designs would allow more experiments in which different types of treatments, and therapists, could be paired with many different types of clients, with many

different specific problems.

Single-case experimental designs highlight the variability in the individual. If a client deteriorates, the reasons for deterioration cannot be speculated upon if only pre and post data are available. It would be much to the advantage of the clinical researchers to have followed the one patient's course during treatment so that the beginning of deterioration could be pinpointed.

Any N=1 study whether empirical or not, experimental or correlational, has limited power in the confirmatory aspect of scientific inquiry. But the same can be said for one isolated nomothetic study (Kiesler, 1981). The external validity of a series of single case designs in similar clients in which the original experiment is directly replicated three or four times can far surpass the experimental group/no treatment control group design (Hersen & Barlow, 1976). "Sophisticated presentation of N=1 research strategy makes it evident that intensive study of the single case involves much more than a single, isolated, N=1 study. Key ingredients include both direct and systematic replications encompassing a series of N=1 studies that address systematically the crucial issues of internal and external validity." (Kiesler, 1981, p. 213). The threats to

internal validity that can be controlled in nomothetic research can be controlled in single case research (Kazdin, 1981). The single case strategy sequentially approximates nomothetic research.

In a discussion of available research designs and methods of analysis applicable to the study of individual subjects Nunnally (1983) suggests the seldom considered possibility to consider each subject as though he or she were a separate experiment and then "glue" subjects together in the context of the experimental design for groups of people. This would be a way of aggregating data on individuals the way Smith, Glass and Miller (1980) and Parloff (1986) aggregate studies to answer questions currently facing psychotherapy research.

The practice of psychotherapy is the application of the scientific method to the single case (Hayes, 1981; Kiesler, 1981). Clinical decision making closely parallels time series methodology. Clinicians "need only (a) take systematic repeated measurements (b) specify their own treatments, (c) recognize the design strategies they are already using, and (d) at times use existing design elements deliberately to improve clinical decision making." (Hayes, 1981, p. 194).

Barlow & Hersen (1973) state that single case experimental designs are particularly well suited for the

study of complex behavior disorders. They review many single case experimental designs that have been employed in in clinical research while providing examples of their use. They believe that the suitability of the designs for clinical research will lead to their increased use.

Single case designs usually begin by observing the client's behavior before treatment. This period is referred to as the Baseline phase. It serves two purposes: 1) to describe the existing level of performance, 2) to predict the level of performance for the immediate future if treatment is not provided. The projection of baseline performance into the future is the implicit criterion against which the treatment is evaluated. If treatment is effective, the actual level of behavior will deviate from the projected level of behavior from baseline performance. After performance stabilizes the treatment can be withdrawn to reassess whether performance under these conditions deviates from the predicted level. "Essentially, data in separate phases of single-case designs provide information about present performance, provide the predicted level of future performance, and test the extent to which prediction of performance from previous phases were accurate." (Kazdin, 1978, p. 630).

Single case experimental designs and the evaluation

of clinical significance parallels the reasoning of the practitioner. Therefore this method of data analysis can be easily adopted by the practitioner who wishes to conduct research. In addition, research presented in the form of clinical significance can be interpreted and used by the practitioner.

Successful clinical practice demands that we use good judgement in choosing optimal treatment for the condition in question (Yeaton & Sechrest, 1981). Practicing clinicians can enhance the quality of their judgment by attending to the strength, integrity, and specific standard of treatment efficacy of a treatment. Strength is the a priori likelihood that the treatment could have its intended outcome. Integrity of a treatment is the degree to which treatment is delivered as intended. Standards of treatment efficacy refer to results aggregated in studies like Smith, Glass and Miller (1980), and the aggregate studies reviewed by Parloff (1986).

Parloff (1986) did an exhaustive review of psychotherapy outcome research between 1980 and 1984. The questions he thinks psychotherapy researchers must answer are: 1) are the positive effects reasonably attributable to psychotherapy or nonspecific placebo effects associated with all therapies 2) can unsafe and inefficient

treatments be identified so a rationale restricting reimbursement can be provided to meet the insurance companies demands, and 3) can the most effective treatments for specific conditions be identified to better serve the patient. Parloff also comments that special problems make implementation of "state of the art" research methodology such as "randomized clinical trials" difficult or impractical. A way to approach the questions and avoid the problem is to use single case experimental designs and clinical significance in data analysis.

The purpose of this study is demonstrate the use of clinical significance in psychotherapy data analysis. In addition, the study is an attempt to support the notion that clinical significance captures an important aspect of improvement that needs to be reported along with statistical significance tests of parameters of distributions in psychotherapy outcome research .

Hypotheses

Hypothesis 1 (Experimental):

Psychotherapy clients showing clinically significant symptom changes will report greater satisfaction and benefit from psychotherapy than (a) a group of clients whose changes fell solely within the statistically significant improvement range, and (b) a group of clients who did not improve statistically or clinically.

Hypothesis 1 (Operational):

Psychotherapy clients showing clinically significant change on the SCL-90-R Global Severity Index will report greater satisfaction and benefit from psychotherapy on the Strupp Post Therapy Client Questionnaire than (a) a group of Clients whose changes fall solely within the statistically significant improvement range on the SCL-90-R, and (b) a group of clients who did not improve statistically or clinically on the SCL-90-R.

Hypothesis 2 (Experimental):

Psychotherapy clients showing statistically significant improvement will report greater satisfaction and benefit from psychotherapy than a group of clients

who did not improve statistically or clinically.

Hypothesis 2 (Operational):

Psychotherapy clients showing statistically significant improvement on the SCL-90-R Global Severity Index will report greater satisfaction and benefit from psychotherapy on the Strupp Post Therapy Client Questionnaire than a group of clients who do not improve statistically or clinically on the SCL-90-R.

Method

Subjects:

Clients: Seventy-five client-therapist dyads were selected for inclusion in the study from a database of 84 therapy cases at the Michigan State University Psychological Clinic. The clients were predominantly working and middle class. All clients agreed to participate in the Clinic's psychotherapy research program. They ranged in age from 16 to 91 years; 68 percent were women.

Therapists: The therapists were graduate students of Michigan State University working at the Psychological Clinic, recruited from the clinic practicum students and interns. They were selected from the database of 84 cases, along with the clients, for inclusion in the study. The therapists range in experience from students in first year practicum to advanced students with several years post-masters degree experience. The predominant theoretical orientation of the therapists was psychodynamic, although other orientations to treatment are represented. Since the study is being conducted after the therapy has been completed the therapists and clients were blind to the hypotheses and purposes of the study.

Materials:

The Symptom Check List 90 Revised

Derogatis' (1983) Symptom Checklist 90 Revised will be used to measure the client's symptoms before and after therapy. The SCL-90-R is a 90 item self administered questionnaire which is composed of nine subscales measuring nine symptom dimensions: Somatization, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism. Subjects are asked the extent to which they are distressed by: 1) headaches 2) nervousness and shakiness inside etc. The subject rates each of the 90 symptom items on a Likert type scale that goes from 0, (not at all) to 4, (extremely). Means are computed for each of the nine subscales. The Global Severity Index (GSI) is the sum of all item responses divided by 90. This represents the best single indicator of the current level of depth of the disorder.

The reliability and validity of the SCL-90-R are discussed in the Administration, Scoring and Procedures Manual (Derogatis, 1983). Reliability is evaluated in terms of internal consistency and test retest reliability. Internal consistency is the consistency with

which the items selected represent each symptom construct. Test retest reliability is the stability of the measure across time.

Internal consistency for the nine subscales was measured by coefficient alpha for a sample of 219 symptomatic volunteers (Derogatis, Rickels & Rock, 1976). Coefficient alpha treats within form correlations among the items as analogous to correlations between alternate forms, and assumes that the average correlations among existing items would be equivalent to the correlation among items in the hypothetical alternate form. The coefficients obtained for this sample were satisfactory and ranged between a low of .77 for psychoticism to a high of .90 for depression.

The test-retest reliability for the SCL90 was checked on a sample of 94 psychiatric outpatients with one week elapsed time between testing. The test-retest reliability coefficients range from .78 for hostility to .90 for phobic anxiety (Derogatis, Rickels & Rock, 1976). Psychopathological symptoms would be expected to be less stable than a characteristic such as intelligence but more stable than "mood". Though psychological symptoms can fluctuate over a period of one week one would not expect much change.

Criterion related validity is supported by several

studies. The SCL-90-R was used in a study evaluating the utility of Research Diagnostic Criteria for predicting differential response to amitriptyline and/or short term interpersonal psychotherapy. The SCL-90-R was found to be sensitive to change and differences in the RDC subtypes (Prusoff, Weissman, Klerman & Rounsaville, 1980). A comprehensive study of the relationship between sexual dysfunction and psychotherapy has utilized the SCL-90-R to demonstrate significant symptom differences between patients assigned to different DSMIII diagnostic categories (Derogatis, Meyer & King, 1981). These studies suggest that the type and severity of symptoms can be assessed using the SCL-90-R.

Construct validity or more specifically concurrent or convergent validity is supported by determining correlation between the scales of the test and other measures of the constructs the scales are intended to measure. Derogatis, Rickels and Rock (1976) compared the dimension scores of the SCL90 with the scale scores from the MMPI. In this study 119 symptomatic volunteers were given the SCL90 and the MMPI. The results of the study were that each dimension of the SCL90 had its highest correlation with a like construct on the MMPI except for the obsessive compulsive dimension for which there is no directly comparable MMPI scale. This study supports the convergent validity of the SCL90.

A similar study of concurrent validity of the SCL90 was conducted by Boleoucky and Horvath (1972). The symptom dimensions of the SCL90 were correlated with those of the Middlesex Hospital Questionnaire (MHQ). The two instruments shared 6 like symptom dimensions. Correlations between like dimensions were computed for a sample of 130 subjects. Correlations ranged from .73 for depression down to .36 for phobic anxiety. For most scales convergent validity is suggested. The global severity index (GSI) and MHQ global correlated .92.

A confirmatory factor analysis (Derogatis & Cleary, 1977) performed on data from 1002 psychiatric outpatients confirmed the hypothesized structure of the SCL-90-R.

The means for the SCL-90-R are available for a sample of 1002 heterogeneous outpatients (Derogatis, 1983). The outpatients came from centers in Johns Hopkins University, the University of Maryland, the University of Pennsylvania and the University of Wisconsin. There were 425 males and 577 females, approximately two thirds white, skewed somewhat towards the lower end of the socioeconomic spectrum. The nonpatient norm group was comprised of 974 individuals, 493 males and 480 females, eight ninths white. Social class data are not available. It represents a stratified random sample from a diverse community in a large eastern state.

The SCL-90-R is a valid and reliable instrument constructed with subjects comparable to the type of subjects in this study. These facts combined with its past use in a related fashion in research such as the Derogatis et al. (1981) study in which symptom changes were evaluated with the SCL-90-R make the SCL-90-R a reasonable choice as a symptom measure for this study.

The Strupp Post Therapy Client Questionnaire

The client's satisfaction with the therapy experience will be measured by the Strupp Post Therapy Client Questionnaire (Strupp, 1969). Fifty-six items such as: "How much have you benefited from therapy?" are evaluated on a Likert type scale from 1, (a great deal) to 9, (not at all).

The original questionnaire had 89 items. Strupp et. al. (1969) administered the questionnaire to clients at the Psychiatric Outpatient Clinic of North Carolina Memorial Hospital, and ended up with 122 completed cases, 59.9% females. The clients ranged in age from 18 to 50, 45.9% were married, and 45.9% were single. They were predominantly middle and working class. Their pretreatment symptoms ranged from loss of interest in life, depression, to interpersonal difficulties, generalized anxiety, and physical symptoms.

The questionnaires were subjected to a cluster analysis. The analysis included: 1) study of response frequencies for each item, 2) intercorrelations (Pearson's "r") among all structured items, 3) systematic studying of statistical relationships, 4) isolation of item clusters, 5) comparison of cluster scores based on items included, and 6) correlations among items and other measures. Step 4, the isolation of cluster items, was conducted by the independent evaluation of members of the project staff. Highly correlated items were grouped to the point at which the staff could no longer agree on the grouping, and correlation among the items dropped to below .50. The analysis produced 10 clusters: 1) Therapist's warmth, 2) amount of change, 3) present adjustment - current status, 4) amount of change apparent to others, 5) therapist's interest, integrity, and respect, 6) (not used) uncertainty about therapist's feelings, 7) intensity of emotional experience, 8) (not used) use of technical terms, 9) degree of disturbance before therapy, 10) therapist's experience/ activity level. The best established clusters were considered to be 1,2,3,4,5, and possibly 9.

The cluster used for this study to measure outcome from the client's subjective perspective was (2) amount of change. This cluster contained items pertaining to:

benefit from therapy, satisfaction with therapy, amount of change, and symptom relief. The inter-item correlations of these items ranged from .91 to .58.

This cluster was used by Lichtenstein (1984) to evaluate psychotherapy outcome from the client's subjective perspective in a study of the effects of client and therapist gender on the outcome and process of psychotherapy. Lichtenstein found intercorrelations among these items ranging from .41 to .73, significant at the .001 level. Eaton (1986) also used this cluster as an outcome measure in a study of therapeutic alliance and outcome.

Since the Strupp post therapy client questionnaire was developed with clients similar to those used in this study, in a context similar to this study, and has been used to measure therapeutic outcome from the client's subjective point of view, it is a reasonable choice for use in this study.

Procedure:

In the Michigan State University Psychotherapy Project database all of the 84 participants have filled out the SCL-90-R prior to beginning therapy and after completing therapy. The participants have also filled out

the Strupp Post Therapy Client Questionnaire after completing therapy. Cases were selected from the database based on change in the level of symptomatic distress after therapy as measured by the pre and post therapy SCL-90-R Global Severity Index (GSI) scores. Three groups of 25 subjects each were created using (GSI) scores and the criteria for statistical and clinical significance operationally defined below.

The Groups:

Group I: Will meet the group criterion of statistical significance and all participants in the group will meet the individual criteria of clinical significance.

Group II: Will meet the group criterion of Statistical Significance, but the participants in the group will not meet the individual criteria of clinical significance.

Group III: Will meet neither the group criterion of Statistical Significance nor will the participants in the group meet the individual criteria of Clinical Significance. This group will be comprised of participants who do not change in therapy.

Operational Definitions of Group Criteria:

Criterion of Statistical Significance: Groups I & II will be considered to meet the condition of statistical significance if traditional between groups hypothesis tests, t-tests between the means of the post therapy SCL-90-R Global Severity Index scores of groups I and III, and groups II and III are statistically significant.

Criterion of Clinical Significance: Clinical significance is evaluated on a participant by participant basis.

Participants in the group will be considered to meet the criterion of clinical significance if each participant in the group meets the following two conditions: 1)

Meaningful Change, and 2) Reliable Change.

- (1) Meaningful Change: A participant will be considered to meet the condition of meaningful change if the post-treatment SCL-90-R (GSI) score is more likely to be drawn from the functional than the dysfunctional distribution. This condition is represented statistically as $X_{post} < c$ where c is defined according to the following formula:
(see table 1)

$$c = \frac{S_0 \bar{X}_1 + S_1 \bar{X}_0}{S_0 + S_1} = \frac{.31(1.39) + .601(.31)}{.31 + .601} = .677$$

- (2) Reliable Change: A participant will be considered to meet the condition of Reliable Change if the Reliable Change Index (RC), defined by the following formula, is greater than 1.96
(see table 1):

$$RC = (X_{post} - X_{pre}) / S_E$$

$$S_E = S_1 \sqrt{1 - r_{xx'}} = .601 \sqrt{1 - .838} = .242$$

Table 1

Symptom Check List 90 Revised Data Used in Determining
Clinical Significance Outcome Criteria

Symbol	Definition	Value
\bar{X}_0	= mean of the SCL-90-R Global Severity Index (GSI) for the well functioning normal population ^A	.31
\bar{X}_1	= pretreatment mean of the SCL-90-R (GSI) for groups I, II, & III combined	1.39
X_{pre}	= pretreatment (GSI) score of a participant	
X_{post}	= posttreatment (GSI) score of a participant	
S_1	= standard deviation of groups I, II, & III combined on the SCL-90-R (GSI) pretreatment	.601
S_0	= standard deviation of the normal population on the SCL-90-R (GSI) ^A	.31
$r_{xx'}$	= test-retest reliability of the SCL-90-R (GSI) ^B	.838
S_e	= standard error of measurement for SCL-90-R (GSI)	.242

^A Based on a nonpatient norm group of 974 individuals from a diverse community in a large eastern state (Derogatis, 1983).

^B Based on a sample of 94 heterogeneous outpatients with one week elapsed between tests (Derogatis, 1983).

All subjects of the 84 meeting the criterion of clinical significance and the criterion of statistical significance, $N = 25$, were included in group I. Of the original 84 subjects 7 were missing the SCL-90-R scores necessary to classify them into a group, so they were dropped from the analysis. Two subjects had actually deteriorated in therapy, the (pre-therapy - post-therapy) difference in GSI scores were $-.85$ and $-.64$. These subjects were dropped from the analysis because deterioration is not consistent with the criterion for membership to any of the groups. The remaining 50 subjects were divided into groups II and III. The 25 subjects with the smallest (pre-therapy - post therapy) SCL-90-R GSI score differences were included in group III, and the other 25 subjects were put in group II.

In accord with the group selection criteria, a t-test between the means of the post therapy SCL-90-R GSI scores of group II ($\bar{M} = .82$; $SD = .337$) and III ($\bar{M} = 1.15$; $SD = .709$) was statistically significant, $t(48) = -2.11$, $p < .04$. A t-test between the means of the post therapy SCL-90-R GSI scores of group I ($\bar{M} = .43$; $SD = .134$) and III ($\bar{M} = 1.15$; $SD = .709$) was statistically significant, $t(48) = -5$ $p < .01$. All subjects in group I met the conditions of clinical significance described above. The mean of the reliable change index (RC) = 2.25

and the standard deviation of the reliable change index $S_{RC} = 2.37$.

A one-way ANOVA comparing these three groups on client satisfaction with therapy using relevant items, (3, 4, 11, and 15) from Strupp's Post Therapy Client Questionnaire as the dependent variable was performed. These groups were compared further using pairwise contrasts.

Results

Hypothesis 1:

The data clearly support hypothesis 1. A group of psychotherapy clients who met the criterion of clinically significant change on the SCL-90-R reported greater satisfaction and benefit from psychotherapy on selected items of the Strupp Post Therapy Client Questionnaire than (a) a group of clients whose changes met only the statistically significant improvement criterion on the SCL-90-R, and (b) a group of clients who did not meet either the criterion of statistical significance or the criterion of clinical significance on the SCL-90-R. A one way analysis of variance comparing a group of clients who met the criteria of statistical and clinical significance (I), to a group of clients who met the criterion of statistical significance (II), and a group of clients who did not meet the criterion of statistical significance or the criterion of clinical significance (III)² was significant $F(2,69) = 4.36, p < .017$ (see table 2).

The significance of the contrast between groups I and III (see Tables 2 and 3), considered with the lack of significance of the contrast between groups II and III indicates that the significant difference in this analysis

is between groups I and III. These contrasts further support hypothesis 1, that clients who have displayed clinically significant symptom change will report the greatest satisfaction and benefit from psychotherapy.

Hypothesis 2:

The results do not support hypothesis 2, that a group of psychotherapy clients who meet the criterion of statistical significance will report greater satisfaction and benefit from therapy than a group of clients who do not meet the criterion of statistical significance or the criterion of clinical significance. The contrast between groups II and III shows a trend toward the support of hypothesis 2, (significance of $p = .099$) when considered in conjunction with the nonsignificant contrast between groups I and II, and the significant contrast when groups I and II are combined and compared to group III.

Post Hoc Analysis:

A post hoc analysis was done comparing pretreatment level of symptomatic distress as measured by the SCL-90-R subscales, across groups I, II, and III, in a series

Note 2 One subject from group II and two subjects from group I dropped out of the analysis because they were missing the post therapy client questionnaire data.

of one-way analysis of variance designs (SCL-90-R symptom scale by group). Four of the 9 subscales were found to be significantly different across groups: interpersonal sensitivity, depression, paranoid ideation, and psychoticism (see table 4).

This indicates that these 4 scales had initial elevations that were high enough for variation in outcome to be possible. This suggests that the sample includes a pretreatment symptom constellation of depression or interpersonal anxiety. These results are also an indication of which symptoms are likely to be alleviated or changed by psychotherapy.

Table 2

The Relationship of Statistical Significance and Clinical
Significance of Level of Symptom Change to Client
Satisfaction.

		Group		
		I	II	III
		Clinical Significance & Statistical Significance	Statistical Significance	No Change
Client	<u>M</u>	2.17	2.64	3.23
Satisfaction	<u>SD</u>	1.2	1.13	1.38
	<u>N</u>	23	24	25

Note: $F(2,69) = 4.36, p < .017$; $MS_{error} = 1.548$

Table 3

Fairwise Contrasts of Groups by Client Satisfaction.

Contrast	T Value	F Value
I vs III	-2.94	.004
I vs II	1.27	.208
II vs III	-1.67	.099
I & II ^a vs III	-2.68	.009

Note: The degrees of freedom were 69 in all contrasts.

^a: Groups I and II were combined and compared to group III.

Table 4

Group Differences on Pre-therapy Symptom Level as Measured
by SCL-90-R Scales.

		Group		
SCL-90-R Scale		I	II	III
Somatization	<u>M</u>	.85	.90	.52
	<u>SD</u>	.57	.74	.62
Obsessive Compulsive	<u>M</u>	1.75	1.72	1.38
	<u>SD</u>	.64	.87	.86
Interpersonal Sensitivity	<u>M</u>	1.73	1.91	1.30 *
	<u>SD</u>	.67	.80	.92
Depression	<u>M</u>	2.2	2.24	1.7 *
	<u>SD</u>	.66	.70	1.02
Anxiety	<u>M</u>	1.82	1.68	1.27
	<u>SD</u>	.89	.84	1.02
Hostility	<u>M</u>	1.13	1.15	1.2
	<u>SD</u>	.65	.82	1.12
Phobic Anxiety	<u>M</u>	.83	.76	.43
	<u>SD</u>	.79	.69	.69
Paranoid Ideation	<u>M</u>	1.19	1.34	.83 *
	<u>SD</u>	.64	.86	.74
Psychoticism	<u>M</u>	1.10	1.04	.66 *
	<u>SD</u>	.51	.70	.63

* $p < .05$; $df = (2,72)$

Discussion

Psychotherapy clients showing clinically significant change on the SCL-90-R reported greater satisfaction and benefit from psychotherapy on selected items of the Strupp Post Therapy Client Questionnaire than (a) a group of clients whose changes fell solely within the statistically significant improvement range on the SCL-90-R, and (b) a group of clients who did not improve statistically or clinically on the SCL-90-R. The data clearly support hypothesis 1. Clinical significance is associated with greater client satisfaction than statistical significance.

Psychotherapy clients showing statistically significant improvement did not report greater satisfaction and benefit from therapy than a group of clients who do not improve statistically or clinically. The results do not support hypothesis 2. Although a trend toward the support of hypothesis 2 was suggested. The lack of confirmation of hypothesis 2 appears to underscore the result with regard to hypothesis 1. Those satisfied with therapy are those who meet the most rigorous criterion for improvement: 1) movement into the normal range of functioning, and 2) change that is reliable and not likely to be due to chance.

The post hoc analysis evaluating pretreatment symptom level by group indicates that there was not a large enough sample of patients describing themselves as compulsives, phobics or with somatic problems to evaluate the effect of clinical significance on satisfaction for these groups. The results are most strongly supported for clients suffering from depression, interpersonal anxiety, paranoia or psychotic symptoms. An alternative explanation is that these four symptom groups represent those manifest symptoms which are most malleable and indicative of therapeutic change for a wide range of disorders.

The post hoc analysis also indicates that outcome was related to pretreatment symptom level. Those with a lower level of symptomatology tended to end up in the no change group and were less satisfied with therapy.

Issues concerning the use of pretreatment and posttreatment level of symptomatology as opposed to the use of change scores to evaluate the outcome of treatment have been discussed by Cronbach and Furby (1970). Cronbach's point that much important information is provided by pretreatment scores is well taken. However, alternative explanations for change are always possible, and the effects of pre and post treatment symptom level can be evaluated independently of change.

The effect of the level of change on client satisfaction is important in itself. Obviously a large symptom change cannot be achieved in an individual with a low level of pathology. A criterion for the evaluation of treatment that identifies those most satisfied with therapy is an important and useful criterion to establish.

Future research should control for pretreatment level of symptomatology. Research comparing a clinically significant (reliable and clinically significant change) group to a group that demonstrated reliable change (more than 2 standard deviations), but not clinically significant change (more likely to be in the well than the dysfunctional range), would be interesting. Research comparing a clinically significant group to a group that represented return to the normal range of functioning, but not reliable change, would also be of value.

Jacobson et al. (1984b) used clinical significance in a study of behavioral marital therapy outcome. The clinically relevant questions of what proportions of couples improve, and how often these improved couples truly remain in the ranks of the nondistressed are addressed. It was found that about a third of the couples actually changed their status from distressed to nondistressed by the end of therapy. In a subsequent study of behavioral marital therapy using clinical

significance to evaluate outcome Jacobson et al. (1985) found similar improvement rates. Though these results may appear more modest than results using the traditional methods of reporting outcome, these results provide a nonambiguous criterion for improvement which provides information on the variability of the outcome data.

From a methodological perspective there are several issues. For those concerned with sample size, the following should be considered: for the purposes of the overall analysis of the three conditions in this study 25 subjects per group was deemed acceptable. Kraemer (1981) has indicated that 20 subjects per group creates sufficient power for most analysis. The cost of adding more subjects is only marginally worthwhile considering the relatively small increase in power more subjects would afford.

This study would be strengthened by multiple measures of the independent variable, (symptom level) and the dependent variable, (client satisfaction) if these measures were correlated and yielded converging results. The advantage of using one well established measure of each construct is that there are no ambiguous results to explain as it is conceivable multiple measures could yield. It should also be noted that this study does not demonstrate a causal relationship between clinically

significant symptom change and client satisfaction, but it does demonstrate that clinical significance is associated with greater client satisfaction.

Another methodological measurement issue is that the definition of clinical significance makes use of the standard error of measurement of the instrument used to measure change. It will be much easier to get meaningful change when an instrument with a small standard error of measurement is used. The impact of instrument selection on clinical significance should be considered when developing a study using clinical significance.

Another methodological issue concerns the use of normative data. In the case of this study the norms of the well functioning normal population on the SCL-90-R GSI were chosen. The question can be raised whether or not it makes sense to compare those with symptoms in the psychotic range with well functioning normals or would it be more appropriate to compare these individuals to nontreated psychotics? Clinically significant improvement of these individuals could be viewed as movement within the severely disturbed (psychotic) range. Although improvement could be viewed as change from the need for institutionalization to being able to live alone.

The issue of the appropriate norm group and the meaningfulness of clinical significance is also an issue

for the "worried well" group who enter therapy relatively symptom free, perhaps to further personal growth or self knowledge. What standards should be used to evaluate these individuals with regard to therapy outcome? Should these individuals be compared against the "idealized fiction" version of normality common in the dynamic perspective of psychology? Several forms of this idealized fiction of normality have been discussed by various psychoanalytic writers such as Jones (1931), Eissler (1960) and Klein (1960). A prototypical definition is given by Levine (1942) as cited in Offer and Sabshin (1966, p. 19). There "normality" is defined in the following manner:

- "1) Nonexistent in a complete form, but existing as relative and quantitative approximation.
- 2) In agreement with statistical averages of specific groups, if that is not contrary to standards of individual health and maturity.
- 3) Physical normality; Absence of physical disease; presence of good structure and function and maturity.
- 4) Intellectual normality.
- 5) Absence of neurotic and psychotic symptoms.
[Levine elaborates later that the normal individual is only relatively free of neurotic and psychotic symptoms.]
- 6) Emotional maturity (especially in contrast with neurotic character formation).
 - a) Ability to be guided by reality rather than fears.
 - b) Use of long-term values.
 - c) Grown up conscience.
 - d) Independence.
 - e) Capacity to "love" someone else but with an enlightened self-interest.
 - f) A reasonable dependence.
 - g) A reasonable aggressiveness.

- h) Healthy defence mechanisms.
- i) Good sexual adjustment with acceptance of own gender.
- j) Good work adjustment."

Does this definition have utility in psychotherapy outcome research or should we be more concerned with controlling symptomatology?

According to the present study, symptom relief appears to be a prerequisite for client satisfaction with therapy outcome.

Client satisfaction has a high level of face validity. It would be difficult to argue for an outcome criterion that could not stand the test of client satisfaction. In this study the no change group was less satisfied with their therapy. Individuals in this group tended to be people with low levels of symptomatology. Unless the goals put forth by Levine

were not achieved by these clients, these results suggest that symptom relief is a preeminent factor in the client's evaluation of therapy and satisfaction with outcome.

Clinical significance is clearly a meaningful way of assessing change in therapy. It is demonstrated to be a reasonable way of defining a client group in the context of outcome research. In addition clinical significance has social importance (Wolf, 1978) in the sense that it has a built in emphasis on behavior change stressed by Kazdin (1980). Clinical significance guarantees that subjects are symptomatically more like normals (Kazdin,

1977, 1978, 1980). Further, when clinical significance is used the variability in the data is made clearly visible rather than being camouflaged in group effects (Jacobson et al., 1984).

Clinical significance will allow statements about the success rate of a treatment. A nonambiguous statement such as 6 out of 20 people improved with treatment A while 12 out of 20 people improved with treatment B can be made. Practitioners need to make choices about which treatment to use with a particular individual (Yeaton & Sechrest, 1981; Parloff, 1986). Clinical significance data can help answer that question. The practitioner can choose the treatment that is effective with most people or make a determination about whether his/her client is more like the 6 people who improved in treatment A or the 12 who improved in treatment B.

As Sargent (1983) found, practitioners prefer studies which report results in terms of clinical significance because it helps them make the decisions that they need to make. Researchers will appreciate clinical significance because it allows a more complete representation of the data that includes variability. Ultimately psychotherapy research results cannot be meaningful unless they are usable by the practitioner.

References

References

- Argas, W. S., Kazdin, A. E., & Wilson, G. T.
(1979). Behavior therapy: Toward an applied clinical science. San Francisco:Freeman.
- Bakan, D. (1966). The test of significance in psychological research. Psychological Bulletin, 66, 423-437.
- Barlow, D. H. (1981). On the relation of clinical research to clinical practice: Current issues, new directions. Journal of consulting and Clinical Psychology, 49, 147-155.
- Barlow, D. H. and Hersen, M. (1973) Single-Case experimental designs. Uses in applied clinical research. Archives of General Psychiatry. 29. 319-325.
- Boleoucky, Z. and Horvath, M. (1972). The SCL-90 rating scale: First experience with the Czech version in healthy male scientific workers. Activitas Nervosa Superior (Praha) 16, 115-116.
- Bolles, R. C. (1962). The difference between statistical hypotheses and scientific hypotheses. Psychological Reports, 11. 639-645.

- Carver, R, P. (1978). The case against statistical significance testing. Harvard Educational Review 3, 378-399.
- Cronbach, L. J., & Furby, L. (1970). How should we measure "change"--Or should we? Psychological Bulletin, 74, 68-80.
- Cronbach, L. J., (1975). Beyond the two disciplines of scientific psychology. American Psychologist. Feb. 116-127.
- Derogatis, L. (1983) SCL-90-R Administration, Scoring & Procedures Manual-II. Towson: Clinical Psychometric Research.
- Derogatis, L., Meyer, J., King, K, (1981). Psychopathology in individuals with sexual dysfunction. American Journal of Psychiatry. 138, 757-763.
- Derogatis, L., Rickels, K., and Rock, A. (1976). The SCL-90 and the MMPI: A step in the validation of a new self-report scale. British Journal of Psychiatry. 128, 280-289.
- Derogatis, L. R., and Cleary, P. (1977) Confirmation of the dimensional structure of the of the SCL-90: A study in construct validation. Journal of Clinical Psychology. 33(4), 981-989.

- Eaton, T., Abeles, N., Gutfreund, M., J. (in press). Therapeutic alliance and outcome: Impact of treatment length and pretreatment symptomatology. Psychotherapy.
- Eissler, K. R., (1960). The efficient soldier. In Muensterburger, W. & Axelrod, S. (Eds.), The psychoanalytic study of society. New York:International Universities Press
- Garfield, S. L. (1981). Evaluating the psychotherapies. Behavior Therapy. 12, 195-307
- Hays, W. L. (1981). Statistics. New York:Holt, Rinehart and Winston.
- Hayes, S. C. (1981). Time series methodology and empirical clinical practice. Journal of Consulting and Clinical Psychology, 49, 193-211.
- Hersen, M., & Barlow, D. H. (1976). Single case experimental designs: Strategies for studying behavior change. New York: Pergamon.
- Hugdahl, K., & Ost, L. (1981). On the difference between statistical and clinical significance. Behavioral Assessment. 3, 289-295.
- Jacobson, N., Follette, W. C., & Revenstorf, D. (1984) Psychotherapy outcome research: Methods of reporting variability and evaluating clinical significance. Behavior Therapy, 15, 336-352.

- Jacobson, N., Follette, W. C., Revenstorf, D., Baucom, D., Hahlweg, K. & Margolin, G. (1984b). Variability in outcome and clinical significance of behavioral marital therapy: A reanalysis of outcome data. Journal of Consulting and Clinical Psychology, 52, 497-504
- Jacobson, N., Follette, W. (1985). Clinical significance of improvement resulting from two behavioral marital therapy components. Behavior Therapy, 16, 249-262.
- Jones, E. The concept of the normal mind. In S. D. Schmalhausen, (Ed.), Our neurotic change. New York: Farrar & Rinehart.
- Kazdin, A. E. (1977). Assessing the clinical or applied importance of behavior change through social validation. Behavior Modification, 1, 427-452.
- Kazdin, A. E. and Wilson, T. (1978). Criteria for evaluating psychotherapy. Archives of General Psychiatry, 35, 407-416.
- Kazdin, A. E. (1978). Methodological and interpretive problems of single-case experimental designs. Journal of Consulting and Clinical Psychology, 46(4), 629-642.
- Kazdin, A. E. (1980). Research Design in Clinical Psychology. New York:Harper and Row.

- Kazdin, A. (1981). Drawing valid inferences from case studies. Journal of Consulting and Clinical Psychology. 49, 193-192
- Kelly, E. L., Goldberg, L. R., Fisk, D. W., & Kilkowski, J. M. (1978). Twenty-five years later. American Psychologist, 33, 746-755.
- Kendall, P. C., & Norton-Ford, J. D. (1982). Therapy outcome research methods. In P. C. Kendall & J. N. Butcher (Eds.), Research methods in clinical psychology. New York: Wiley.
- Kiesler, D. J. (1981). Empirical clinical psychology: Myth or reality? Journal of Consulting and Clinical Psychology. 49, 212-215
- Klein, M. (1969). On mental health. British Journal of Medical Psychology. 33, 237-241.
- Kraemer, H. C., (1981). Coping strategies in psychiatric clinical research. Journal of Consulting and clinical psychology. 49(3), 309-319.
- Levine, M. (1942). Psychotherapy in medical practice. New York: Macmillan.
- Lichtenstein, A. B. (1984). The effect of client and therapist gender on the outcome and process of psychotherapy. Unpublished Dissertation, Michigan State University.

- Lick, J. (1973). Statistical vs. clinical significance in research on the outcome of psychotherapy. International Journal of Mental Health. 2, 26-37.
- Maletzky, B. M. (1981). Clinical relevance and clinical research. Behavioral Assessment. 3, 283-288
- Minkin, N., Braukmann, L. J., Minkin, B. L., Timbers, G. D., Timbers, B. J., Fixsen, D. L., Phillips, E. L., and Wolf, M. M. (1976). The social validation and training of conversational skills. Journal of Applied Behavior Analysis. 9, 127-139.
- Nunnally, J. and Kotsch, W. (1983). Studies of individual subjects: Logic and methods of analysis. British Journal of Clinical Psychology. 22, 83-93.
- Offer, D., & Sabshin, M. (1974). Normality. New York: Basic Books.
- Parloff, M., London, P. & Wolf, B. (1986). Individual psychotherapy and behavior change. Annual Review of Psychology. 37, 321-349
- Patterson, G. R. (1974) Intervention for boys with conduct problems: Multiple settings, treatments, and criteria. Journal of Consulting and Clinical Psychology. 42. 471-481.

- Prusoff, B., Weissman, M., Klerman, G. L., and Rounsaville, B. J. (1980). Research diagnostic criteria subtypes of depression: Their role as predictors of differential response to psychotherapy and drug treatment. Archives of General Psychiatry. 37, 796-801.
- Sargent, M. & Cohen, L. N. (1983). Influence of psychotherapy research on clinical practice: an experimental survey. Journal of Consulting and Clinical Psychology. 51(5), 718-720.
- Smith, M. S., Glass, G. V., & Miller, T. L. (1980). The benefits of psychotherapy. Baltimore: Johns Hopkins University Press.
- Strupp, H. H., Fox, R., Lessler, K. (1969). Patients view their psychotherapy. Baltimore: The Johns Hopkins Press.
- Strupp, H. H. and Hadley, S. W. (1977). A tripartite model of mental health and therapeutic outcomes with special reference to negative effects in psychotherapy. American Psychologist. March 187-195.
- Strupp, H. H. (1981). Clinical research, practice, and the crisis of confidence. Journal of Consulting and Clinical Psychology. 49, 216-219

- Winch, R. F. and Campbell, D. T. (1969). Proof? no. evidence? yes. the significance of tests of significance. The American Sociologist. 4. 140-143.
- Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. Journal of Applied Behavior Analysis 11, 203-214.
- Yeaton W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. Journal of Consulting and Clinical Psychology. 49, 156-167.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03082 2401