This is to certify that the

thesis entitled

THE CONSTRUCT VALIDITY OF A SELF-REPORT INSTRUMENT IN MENTAL HEALTH EVALUATION

presented by

Lorraine LaFerriere

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Psychology

Major professor

Date August 7, 1979

O-7639

THE CONSTRUCT VALIDITY OF A SELF-REPORT INSTRUMENT IN MENTAL HEALTH EVALUATION

By

Lorraine LaFerriere

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Psychology

1979

ABSTRACT

THE CONSTRUCT VALIDITY OF A SELF-REPORT INSTRUMENT IN MENTAL HEALTH EVALUATION

By

Lorraine LaFerriere

Research in mental health evaluation has evolved primarily during the last twenty years, with attention focused on issues of methodology and design (Campbell and Stanley, 1963; Kenny, 1975) and more recently on issues of measurement (Waskow and Parloff, 1975; Erickson, 1975; Ellsworth, 1975). An emerging consensus suggests that evaluating mental health programs requires more than traditional length-of-stay and rehospitalization measures; measures of how the patient is functioning socially and psychologically are necessary to determine whether patients truly benefit from treatment. Less consensus appears around what measures should be used, and whether they should be based on the client's own report of his functioning, or a clinician's or significant other's report.

This study is an examination of the validity of a client's self-report instrument in mental health evaluation. The study was part of a larger effort to pilot test a set of procedures and instruments in outcome evaluation, jointly sponsored by the Michigan Department of Mental Health and the Urban Institute. In 1977-1978,

the pilot was implemented in five community outpatient units and two inpatient units. Several hundred outpatients and inpatients were asked to complete pre- and post-treatment questionnaires about their psychological and social functioning, and to name a "significant other" who would complete similar questionnaires. Clinicians rated the clients' symptomatic disturbance at intake. A random sample of several hundred community residents also completed the self-report instrument.

A substantial portion of the eligible populations, especially inpatients, did not complete questionnaires at intake and/or at follow-up. Data loss analysis indicates the subjects completing questionnaires were less disturbed and were of a higher socioeconomic status than subjects not completing questionnaires. Thus, the generalizability of the findings are limited to the healthier portion of inpatients and outpatients.

The primary instrument in this study, the Brief Symptom Inventory (BSI), was selected for its ease of administration, potential applicability for outpatients as well as inpatients, and the well established psychometric properties of its parent instrument, the SCL-90 (Derogatis, 1977). In examining the construct validity of the BSI for evaluation purposes, three criteria were established --whether the scores (a) differentiated between inpatients, outpatients and community residents; (b) were congruent with other measures of symptom disturbance; and (c) were sensitive to pre- and post-treatment changes in symptom disturbance.

Results of the study indicate the BSI has construct validity for women, but not for men. Most of the validity criteria were satisfied for female outpatients and inpatients. By contrast, male inpatients did not report significantly more symptoms than male outpatients, nor did either sample of men report a lessening of symptoms after treatment. The correspondence between the BSI and reports of symptoms by clinicians and significant others was greater for outpatients than for inpatients, with the lowest correspondence being for male inpatients.

Investigation of the validity of the BSI is confounded with other possible factors, including possible inherent sex differences in the experience and/or reporting of symptoms; methodological questions regarding measurement of change; and some evidence that symptomatic disturbance may be more salient factors for women seeking and benefitting from treatment than men. Further research is necessary to examine these issues in more depth.

The evidence from this study suggests that while the BSI may be a valid instrument for use with outpatients, it is not valid for use with inpatients. The construct validity of this scale is strongest for female outpatients, and weakest for male inpatients. The combined results of a high percentage of data loss and the lack of instrument validity for males render this self-report model inappropriate for evaluating the effectiveness of mental health services.

ACKNOWLEDGMENTS

I would like to acknowledge and thank the following people who contributed to the successful completion of this study:

First, I would like to thank the staff members in the Clinton-Eaton-Ingham Community Mental Health Center, the St. Lawrence Inpatient Unit, and the Michigan Institute of Mental Health. They gave many hours out of their busy schedules to gather data and have clients complete questionnaires. The willingness of several hundred clients from these agencies to complete questionnaires is also appreciated.

I owe the greatest thanks to my staff on the project. Cathy Comins worked with me from the beginning of the study, and handled a wide range of tasks with great competence and sensitivity. Judy Pfaff was responsible for most of the computer analyses in the project, which she completed with much patience and skill. Both Judy's and Cathy's support, good humor and friendship throughout the project will always be remembered.

This study was made possible by the Budget and Evaluation Section of the Michigan Department of Mental Health. I sincerely thank Ron Uken, Administrator of this section, for his support of this project and mental health evaluation in general. I would also like to thank Rick Spates, Susan Lawther, Sandy Herman and Stu

Halgren for the many hours they spent discussing the project and assisting with the data analyses.

I would also like to thank Harry Hatry and Al Schainblatt from the Urban Institute. Their original proposal for the study, and their vigilance in implementing and critiquing the project were important to its successful development.

My appreciation is also extended to the chairman of my dissertation committee, Don Grummon, and its members, Bertram Stofflemayr, Bob Zucker, and Bill Crano. I am grateful to Don and the other members for their sustained interest in the study, and for their helpful suggestions and support during its two years of fruition. Although not members of my dissertation committee, I would like to thank Bob Calsyn for being my first and best teacher at MSU of methodological issues in evaluation research, and Bob Ellsworth for being so supportive and informative about these issues as well.

Finally, I would like to thank my husband, Barry Wright. He gave me the greatest support possible by encouraging me during the discouraging times, and by being willing to think through the difficult parts of the study. His intellect, wit and sense of the absurdity of it all was invaluable to me.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
INTRODUCTION	1
LITERATURE REVIEW	6
The Brief Symptom Inventory	7
History and Development	7
Psychometric Properties	9
Factor Structure of the SCL-90	12
Construct Validation	13
Designing an Evaluation System	16
Instrument Selection	16
Selecting an Informant	19
Selecting a Follow-Up Period	23
Data Loss	24
Methodology	25
Previous Outcome Studies	29
METHOD	36
Description of the Study	36
Subjects	37
Instruments	39
Data Collection	42
Hypotheses	43
RESULTS	47
Subject Participation Rates	47
Representativeness of Intake Sample	49
Representativeness of Follow-Up Sample	53
Validity: Differentiation of Groups	60
Validity: Concurrence Between Measures	64
Sensitivity of BSI to Change	72

Page

DISCUSSION 79

Generalizability of Results. 79
Tests of Construct Validity. 80
Alternative Theoretical Explanations 84
Interpretation of Sex Differences 84
Measuring Change. 87
Treatment Effectiveness 89
Concurrence Among Informants. 89
The Evaluation System 91
Recommendations for Further Research 94
Conclusions. 95

BIBLIOGRAPHY 97

APPENDICES 103

A PARS Scores of Outpatients and Inpatients at Intake . . 104

B Client Satisfaction with Service. 106

C Consent Forms and Instruments 108

LIST OF TABLES

Table		Page
1.	Demographic Characteristics of Community Residents and Clients in Tri-County Study.	38
2.	Number of Subjects With BSI, PARS and HPRS Questionnaires	44
3.	Demographic Characteristics of Subjects Completing and Not Completing an Intake Questionnaire	50
4.	HPRS Scores of Subjects Completing and Not Completing A BSI at Intake.	52
5.	Characteristics of Outpatients Returning and Not Returning Follow-Up Questionnaires	54
6.	Characteristics of Inpatients Returning and Not Returning Follow-Up Questionnaires	56
7.	Effects of Data Loss on Tri-County Study.	59
8.	Brief Symptom Inventory Intake Scores for Inpatients and Outpatients.	61
9.	Brief Symptom Inventory Mean Intake Scores for Inpatients, Outpatients and Community Residents.	62
10.	Correlations Between Inpatient Scores on Symptomatic Disturbance Scales on the BSI and PARS	65
11.	Correlations Between Outpatient Scores on Symptomatic Disturbance Scales on the BSI and PARS	67
12.	Mean Scores and Correlations Between Corresponding HPRS and BSI Scales.	70
13.	ANOVA of BSI Scores--At Intake, Three and Six Months.	75

INTRODUCTION

State and community mental health treatment programs are serving an increasing number of people experiencing a variety of emotional and behavioral problems. In order to provide these services, state and federal agencies have been increasing the funds available to these treatment programs, with the expectation that effective services will be provided to the treatment recipients. Many of these agencies, however, are not able to measure the effectiveness of their programs. Objective and quantifiable evidence of the impact of treatment is not routinely available to persons administering, funding and evaluating the programs.

A well-designed evaluation system can provide information for a variety of decisions. At a minimum, the type and degree of disturbances and problems characteristic of the population served can be known. Hypotheses regarding the impact of certain types of treatment for certain types of client problems can be generated. Successive tests of these hypotheses will bring the evaluator closer to measuring the "true" effectiveness of the treatment program. This information can help identify which programs appear to be successful and not successful in treating certain types of clients. When the evaluator has identified those program variables influencing the less desirable outcomes, the programs can be re-designed to improve an evaluation

system. If the instruments are not appropriate for the population being assessed, if their psychometric properties are questionable, or if completing them is too time consuming or confusing, the entire evaluation system is in jeopardy. The selection of the instruments is one of the most important steps in designing an evaluation system.

Current research in outcome evaluation (Waskow and Parloff, 1975; Erickson, 1975; Ellsworth, 1975) suggests that outcomes should be measured on a variety of dimensions, primarily including the social and psychological functioning of the clients. No one single instrument has emerged as a "best" measure for outcome evaluation; often, the use of several measures completed by different informants familiar with the client's functioning is ideally recommended.

The recommended composite of evaluation instruments typically includes a report from the client about his or her emotional state. Client self-report instruments, especially symptom checklists, are seen as easy to complete and score, and as allowing clients to report systematically on a variety of psychiatric problems (Derogatis, 1974; Schainblatt, 1978). Questions remain, however, as to which instruments are most appropriate for what types of clients; research efforts to date have not systematically addressed this issue.

Given the lack of consensus regarding which instruments are most useful in measuring the effectiveness of mental health programs, the instruments selected for a system should be pilot tested and then assessed according to criteria relevant to their intended use. Unless this step precedes the actual utilization of the evaluation data, the

evaluator will not be able to determine whether the results reflect the strengths or weaknesses of the instrument or the actual outcome of treatment.

The usual approach to assessing an instrument is to investigate its psychometric properties. If reasonably high reliabilities are established for an instrument and it meets some validity criteria, the instrument becomes commercially available and may be used for a variety of purposes. Its use may or may not relate to the original purpose(s) for which it was designed. This approach to selecting an instrument is not sufficient to insure the instrument is used in an appropriate context.

This study is an assessment of the validity of an instrument selected for use in a pilot project evaluating mental health treatment programs. The instrument, the Brief Symptom Inventory (BSI) (Derogatis, 1977) is a 53-item questionnaire measuring the self report of the respondent's psychological symptoms. The BSI was selected for this study for several reasons. First, it incorporates many of the advantages of its parent instruments, the SCL-90 (Derogatis, 1977) and the HSCL (Derogatis et al., 1974), which have established reliability and validity properties, and have been used extensively in psychological assessment research. The BSI has the advantage of being shorter than the SCL-90 and includes more symptoms manifested by both outpatients and inpatients than the HSCL. Finally, the BSI is an instrument similar to those being used or recommended by many state and federal agencies developing mental health evaluation systems. As such, it is representative of an approach to mental health evaluation

that could be incorporated into the evaluation of hundreds of publicly funded mental health programs.

In assessing the BSI, this study addresses several questions relevant to establishing the instrument's construct validity. Does the BSI accurately assess the client's psychological symptoms at intake and follow-up? Does it differentiate among groups of people with inpatient, outpatient or "normal" status? Is the symptomatic disturbance measured by the BSI congruent with the clinician and a significant other's report of the patient's symptomatic distrubance? Does the BSI reflect change that would be predicted on the basis of clinical experience and prior research?

This study is one step in assessing an evaluation system being pilot tested by the Michigan Department of Mental Health. Other aspects of the pilot--the social functioning and client satisfaction measures, the feasibility and cost of administering the system and the efficiency of the procedures--are being assessed in separate studies. The evaluation model used in this pilot was tested in outpatient and inpatient units in Clinton, Eaton and Ingham counties to determine whether it is an appropriate model to implement on a statewide basis. Several hundred inpatients, outpatients and non-patients were tested at pre- and post-treatment intervals. The system incorporated many of the design features recommended in national reports of evaluation strategies; as such, it parallels several other states' efforts to develop statewide evaluation systems.

The results of this inquiry will further refine the information available to evaluators regarding the selection of appropriate

instruments for evaluating treatment programs. It will explore the usefulness of the instrument with regard to particular aspects of the evaluation system and identify the types of patients and symptom disturbance measurable by this instrument. The large sample of out-patients, inpatients and non-patients available to test these issues, as well as the multiplicity of instrumentation in this study, should provide uniquely valuable information in the design of an evaluation system.

LITERATURE REVIEW

The literature relevant to this study is drawn from several different areas of inquiry. In order to test one aspect of an evaluation system, the design of the entire system must be theoretically and practically sound. The validity of the BSI in measuring treatment outcome is intricately related to the other aspects of the system--in particular, the selection of a multiple core battery of instruments and the methodology used to assess the impact of treatment.

The results of previous psychotherapy outcome research are also relevant to this project. An awareness of outcomes of other studies assessing the effectiveness of both outpatient and inpatient programs places this study in better context. If one assumes treatment programs help some people and do not help others, the question becomes, "How do you design a system which measures that change?" Reviews of previous studies measuring treatment effectiveness are helpful in the formulation of hypotheses and subsequent analysis of the results of this study.

The literature review begins with a discussion of the Brief Symptom Inventory, its history, development, and psychometric properties. The approach to establishing the validity of the BSI is then discussed. Following the discussion of the validation of the BSI is a review of literature relevant to designing an evaluation system,

particularly selecting the instruments and controversial methodological issues in measuring change. Finally, the results of other outcome studies, including psychotherapy research and inpatient treatment programs, are reviewed.

The Brief Symptom Inventory

History and Development

The Brief Symptom Inventory has developed over a number of years in conjunction with efforts to establish a reliable and valid method of assessing psychological symptomatology through a patient self-report mechanism. These efforts can be traced to the Cornell Medical Index (Wider, 1948) and the "Discomfort Scale" (Parloff, 1953). The Clinical Psychometrics Research Unit at John Hopkins has extended this work into the development of the Hopkins Symptom Checklist and the SCL-90-R (Derogatis, 1971, 1972, 1974). Each of these instruments has refined and improved the self-report of psychological symptoms as an assessment tool.

The Brief Symptom Inventory (BSI) is a brief form of the SCL-90-R. The subject is instructed to indicate on a five-point scale how much he was bothered by a list of 53 symptoms during the last week. The responses range from "not at all" to "a great deal." The 53 items on the BSI are those items which had the highest factor loadings on the nine primary symptoms scales of the SCL-90-R. The BSI takes approximately ten minutes to complete, and was designed for use with the debilitated patient who cannot complete the 90-item SCL-90-R and for use in a mental health clinic where the number of

clients seeking treatment precludes a more lengthy psychological assessment.

The BSI measures symptomatic disturbance along the same nine primary dimensions and three global indices as the SCL-90-R. The primary dimensions are: Somatization, Obsessive-Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, Psychoticism. These symptom dimension scores provide a profile of the nature and intensity of the patient's symptoms along standard, clinically recognizable classifications of psychopathology. The scores also provide discrete information about significant details of the patient's condition, including his suicidal tendencies, experience of guilt, and aggressive or violent thoughts.

Three global indices of symptomatic disturbance are also derived in scoring the BSI. The General Severity Index (GSI) is the average symptom score and is the best single indicator of the disturbance level. The Positive Symptom Total (PST) equals the number of symptoms endorsed. The Positive Symptom Distress Index (PSDI) measures the intensity of symptom distress when all zero responses (no disturbance) are eliminated.

Derogatis (1977) reports several advantages to using a psychological self-report instrument. The primary advantage is its reflection of the patient's experience of psychological distress. This subjective report of the patient's perception of his inner-state is unique and valuable information for the clinician and researcher. Other advantages of the self-report symptom checklist include ease

of administration, the minimal professional time required for assessment, and actuarial methods of scoring and interpretation.

The self-report symptom inventory has a variety of potential uses. The SCL-90 has been used as a screening instrument to determine need for mental health service (Craig and Abeloff, 1974) and as a standard source of information regarding the clinical status of a patient. Derogatis also reports that the self-report inventory has become a frequent means of operationally defining "normality" versus "abnormality" (Derogatis, 1977).

Psychometric Properties

Although the reliability and validity of the BSI has not been formally tested, these properties of the SCL-90 have been established. Because the BSI is a subset of the most critical items in the SCL-90-R, the psychometric testing of the parent instrument is relevant to establishing reliability and validity properties of the BSI. This relevancy is supported by evidence of correlations ranging from .92 to .98 on the similar symptom dimension scores of the BSI and SCL-90-R (Derogatis, 1977).

Both the internal consistency and test-retest reliability of the SCL-90 for outpatients have been established with coefficients ranging from .77 to .90 (Derogatis, Rickels and Rock, 1976; Derogatis, 1977; Edwards, 1977). The high internal consistency coefficients indicate the homogeneity with which the items selected to represent each symptom construct actually reflect the underlying factor. The stability of the report over time was established by the high correlations between the reports of 90 outpatients assessed twice during

a one-week interval. A recent study assessing the reliability of five self-report symptom inventories for a nonpatient and outpatient sample concluded that the SCL-90 approaches perfect reliability in terms of internal consistency and test-retest reliability (Edwards, 1977).

Several studies have explored different aspects of the validity of the SCL-90. A major study of the concurrent validity compared scales of the "90" with similar scales of the MMPI (Derogatis, Rickels and Rock, 1976). MMPI clinical scales, as well as Wiggins content scales and Tryon cluster scales correlated highly (range from .40 to .75 with a sample of 219 symptomatic volunteers) with comparable SCL-90 scales. Similarly high correlations were found in a concurrent validation study of the "90" with the Middlesex Hospital Questionnaire (Bolelоucky and Horvath, 1974).

The discriminative validity of the "90" has been investigated in two studies of medical patients judged to be likely to manifest psychological symptoms. Both Craig and Abeloff (1974) and Abeloff and Derogatis (1977) found that cancer patient's symptom profiles on the "90" were similar to these of psychiatric patients. A study of clinical depression among participants in a methadone maintenance program found the scales of the "90" to discriminate between the depressed and non-depressed groups (Weissman, et al., 1976).

Construct validity of the "90" was investigated to determine the correlation between the operations of measurement and the theoretical constructs which they purport to measure. Derogatis and Cleary (1977) completed a study in which factor scores on the nine

primary dimensions (sample of 1,000 outpatients) were rotated and compared with hypothesized structure of the "90". The match between the theoretical and empirical structures was quite high, thus establishing supportive evidence for the construct validity of the SCL-90-R.

Unfortunately, the reports on the validity and reliability of the "90" have been limited primarily to studies using outpatient and nonpatient samples. One report (Derogatis, 1977) found SCL-90 scores gathered on 200 inpatients were two standard deviations above the nonpatient sample, but did not systematically assess the discriminant validity of the "90" with inpatients. The validity of the "90" in measuring treatment effectiveness also has not been established. Although Derogatis (1977) reported that inpatient scores fell about one standard deviation (measured against community norms) from admission to discharge, similar reports on change scores for outpatients have not been reported.

These studies indicate substantial reliability and validity of the SCL-90 in assessing outpatients' and nonpatients' self-report of symptomatic disturbance. Derogatis' initial reports suggest the SCL-90 is valid for inpatients, but this has not clearly been established, particularly as an instrument sensitive to pre- and post-treatment change in symptomatic disturbance. Because these investigations of the psychometric properties of the "90" serve as a proxy measure for the reliability and validity of the BSI, they bear on this study's construct validation of the BSI. Particular attention must be paid to whether the BSI is valid for inpatients, and is

sensitive to pre- and post-treatment differences in the inpatient and outpatient populations.

Factor Structure of the SCL-90

Investigations of the factor structure of the SCL-90 began with factor analysis of the original 58-item HSCL (Derogatis, Lipman, Covi, Richels, 1971, 1972). These studies revealed that depression and anxiety factors accounted for the bulk of the variance, and had factorial invariance across social class and psychiatric diagnosis.

The SCL-90 was created by the addition of 32 items concerned with symptoms of more serious psychopathology to the HSCL. Derogatis and Cleary (1977) found the "90" to have factorial invariance across sex on all nine of its primary symptom dimensions. Lipman, Covi and Shaprio (1977) identified eight similar factors in a study involving chemotherapy of depressed patients.

A recent study by Hoffman and Overall (1978) was prompted by their concern that previous factor studies of the "90" had been conducted on patients selected for symptoms of anxiety and/or depression. Using a clinic population unselected for diagnosis and representative of an outpatient population, they found the rotated factors of the "90" to be quite similar to the responses of anxious or depressed patients. They found the Depression, Somatization and Phobic Anxiety factors to be the most clearly defined, and most consistent with other work on the HSCL and SCL-90. However, the large portion of the variance accounted for by the first unrotated factor and the high intercorrelations among factors suggest that the "90" measures more

of a general complaint or general discomfort dimension than distinct dimensions of psychopathology. Therefore, they recommend limiting the use of the SCL-90 to its measurement of global disturbance level, rather than individual symptom scales.

Investigation of the factor structure of the SCL-90 has direct bearing on the present study of the validity of the BSI. Because no studies of the factor structure of the BSI have been reported, the structure must be assumed to be no better defined than that of the "90". Therefore, the global measures of disturbance on the BSI will be the primary focus of this analysis. Symptom scale scores will be reported for information purposes, but their validity will not be considered in-depth.

Construct Validation

The validity of the BSI must be established with respect to the purpose for which the BSI is being used. As Nunnally (1967) described in his basic text on psychometric theory, "a measuring instrument is valid if it does what it is intended to do. . .one validates not a measuring instrument, but rather some use to which the instrument is put" (p. 75). Cronbach (1970) reiterated this practical approach to validation in suggesting the essential question is "How valid is the interpretation I propose for this test?"

Several different types of validity are described by Cronbach (1970), each of which asks a different question:

(1) Criterion validity: how do measures of some valued performance (criterion) relate to the test score?

(2) Content validity: do the observations truly sample the universe of tasks or the situations they are claimed to represent?

(3) Construct validity: how can scores on the test be explained psychologically? Does the test measure what it is said to measure?

Whereas criterion-validity is often established through single examinations of the correlation between test scores and certain outcome criteria (e.g., later performance on a job), construct validity is established through a long-continued interplay between observation, reasoning and imagination.

Construct validity is described by Anastasi (1968) as a comprehensive concept of validity that includes content and criterion-oriented validity; it is a broad, enduring and abstract kind of behavioral description and, as such, requires the gradual accumulation of information about the test from a variety of sources. The following description of construct validity by Anastasi (1968) is applicable to the current investigation of the BSI as a tool for evaluating mental health services:

> The theoretical construct measured by a test can be defined in terms of the operations performed in establishing the validity of the test. Such a definition would take into account the variables with which the test correlated significantly, as well as the conditions found to affect its scores and the groups that differ significantly in such scores. . . It is only through the empirical investigation of the relationships of test scores to other external data that we can discover what a test measures (p. 122).

Three common methods of establishing construct validity cited by Magnusson (1966) are also highly applicable to the BSI:

1. The study of differences between groups which should differ according to the theory for the variable;

2. The study of how the test results are influenced by changes in individuals or environment which, according to the theory, should respectively influence or fail to influence the individual's positions on the continuum; and,

3. The correlation between different tests which are assumed to measure the same variable (p. 131).

These three aspects of establishing construct validity are utilized in the present investigation of the validity of the BSI. Using Magnusson's criteria to determine whether the BSI is a valid instrument for assessing pre- and post-treatment symptomatic disturbance of outpatients and inpatients, the BSI should:

1. Discriminate among inpatients, outpatients and nonpatients who are likely to have different levels of symptomatic disturbance;
2. Be sensitive to pre- and post-treatment differences in symptoms of outpatients and inpatients; and,
3. Bear a direct relationship to other measures of symptomatic disturbance.

Thus, the construct validity of the BSI will be assessed by testing a series of hypotheses about symptom disturbance of recipients and non-recipients of mental health services, rather than according to a single measure of correlations between test scores and criterion scores. The hypotheses described above are further developed in reviewing research findings regarding pre- and post-treatment clients' reports of symptoms and their relationship to other measures of symptom disturbance. Methodological issues in

evaluation research, as well as the results of psychotherapy outcome research also provide quidelines for the current investigation.

Designing an Evaluation System

The issues inherent in evaluating the outcome of mental health treatment programs have received increased attention in the psychological literature in the past several years. An increasing number of academics are addressing themselves to the issues of how to design a system that is sensitive to the objectives of the treatment program and also produces valuable information to the planners and policy-makers for mental health programs.

Instrument Selection

Several authors have recently discussed the need for development of more relevant outcome measures for evaluating hospital treatment programs than traditional statistics measuring length of treatment and return to treatment or relapse (Erickson, 1975; Speer, 1976; Lick, 1973). The objective of mental health treatment programs is not to treat people for a short or long period of time, or to prevent them from returning to treatment at a later time. Measures must be used which report whether the patient's psychological and behavioral functioning improved as a result of treatment.

As Erickson (1975, p. 526) suggested, outcome measures should reflect the objectives of the treatment program:

> In summary, it appears that patient movement statistics are so full of fallacies and are so difficult to interpret meaningfully, especially in brief-stay settings, that they must be regarded as useless or misleading if taken by themselves. Efficiency and effectiveness depend on the time

> needed to reach goals. The optimal treatment period depends on the needs of the patient and the resources of the program in question. At this point, patient movement statistics are as dated as bed census statistics as a basis for computing cost effectiveness. We can no longer bypass the direct assessment of improvement in the psychosocial functioning of patients admitted to our care.

A report by the National Institute of Mental Health recently recommended that outcomes of mental health programs be measured with a battery of instruments focused primarily on client distress and social functioning in the community (Waskow and Parloff, 1975). A group of NIMH consultants noted for their contributions to psychotherapy research were asked to select a core battery of instruments. In doing so, they stressed the importance of selecting instruments that could be used for comparing outcomes across a variety of treatment approaches and with a variety of patient populations. Included in their recommendations for outcome measures were the Hopkins Symptom Checklist (HSCL) (Derogatis, et al., 1974) and the Personal Adjustment and Role Skills Scales (PARS) (Ellsworth, 1975). Regarding the HSCL, Waskow said "Clearly a battery must include a measure of symptomatology from the perspective of the patient" (p. 253); he went on to point out that the PARS offers "an extremely important perspective to tap, especially if one is interested in the community adjustment and in the overt social behavior of the patient" (p. 258).

The HSCL is a symptom checklist very similar to the BSI, the major difference being that the BSI measures symptoms in a wider range of areas (e.g. the BSI includes paranoid and psychotic symptoms). This difference makes the BSI more relevant for an inpatient population; the NIMH report was concerned primarily with measuring

outcomes for an outpatient population. In a comparable vein, the report suggested the PARS might be more appropriate for former inpatients than for an outpatient neurotic population. Ellsworth (1975) reported, however, wide use of the PARS for both outpatient and inpatient populations. Thus, the selection of the BSI and PARS as core instruments in this study is congruent with the NIHM recommendations, although somewhat adjusted for use with an inpatient and outpatient population.

Measures of client satisfaction with service have also been recommended (McPhee, 1975) as an essential part of an evaluation system. These measures provide information about the client's perception of the impact of treatment on their problem, and a general measure of their overall satisfaction with the treatment program. In addition, assessing the family burden arising from the patient's illness has been included in recommended evaluation instruments (Arnoff, 1975). As an increasing number of patients are being treated and maintained in the community rather than hospitals, it is necessary to assess the impact of this shift on the patient's family.

Measures of patient satisfaction with service and of family burden are included in the present study. The full complement of outcome measures, including symptomatic disturbance, role functioning, satisfaction with service and family burden reflects the complexity of treatment objectives of mental health programs. Although these measures may not inter-correlate in a way that offers a simple definition of treatment success, the support for using a multi-

variate outcome measure is convincing (Erickson, 1975; Waskow and Parloff, 1975; Speer and Tapp, 1976).

Selecting an Informant

The choice of informants, the person providing the information about the patient, is an important factor in the design of the evaluation system. This choice is often intertwined with, or imbedded in, the selection of the instrument, as the instrument is designed to measure a certain person' perspective on the patient's functioning. There are arguments for and against the choice of the patient, therapist, significant other (patient's friend or relative), or research interviewer as the informant. Empirical analysis of these alternatives offers some guidance, but the ultimate choice is more determined by the question "Who wants to know what information at what cost and for what purpose?" (Maguire, 1977).

The client may be the best provider of information regarding his subjective state and feelings of distress or well being, but the validity and reliability of their reports of behavioral adjustment are questionable (Ellsworth, 1968; Paul, 1966; Carr and Wittenbaugh, 1969). Advantages of having the client provide information regarding his subjective discomfort include the low expense in obtaining this information (especially in a brief, self-report measure) and its relevancy in determining whether the client felt better as a result of treatment (Schainblatt, 1977). Disadvantages include the possible distortion of reporting due to a patient being too distressed to complete the questionnaire or excessively defending against

acknowledging his problems. Others argue that patients seeking treatment should not be burdened or alienated by having to complete a questionnaire that is designed to evaluate the agency rather than directly assist the patient (Maguire, 1977).

In traditional settings, the therapist is regarded as the most qualified to evaluate the outcome of treatment. Ideally, the therapist is the most sensitive and objective observer of the patient and the extent to which he/she improves over the course of treatment. Therapists' ratings of improvement have frequently been used as measures of outcome, as evidenced by their use in 65 percent of the 165 studies reviewed by Luborsky, et al., (1971).

Arguments against therapists' ratings include the high cost of using professional time to evaluate treatment, especially in public mental health agencies where therapists are excessively burdened with record-keeping related to accountability and evaluation requirements (Maguire, 1977). A strong bias against using therapists' ratings also exists among administrators who believe therapists will inflate their ratings in order to present a better picture of their efforts. This problem is highlighted in Garfield et al.'s report (1971) that therapists' ratings show more improvement than pre- and post-difference scores on client self-report inventories, such as the MMPI. Finally, a strong argument against using therapists' ratings at discharge was raised by Ellsworth (1968) when he found them to be unrelated to measures of post-hospital adjustment.

Significant other ratings are being used more frequently as their reliability and validity has been demonstrated, particularly

in assessing patient adjustment in the community (Ellsworth, et al., 1968). Depending on the data collection procedures, this approach can be less expensive and burdensome than collecting the same information from patients and/or therapists. Disadvantages of this approach center primarily around the high data loss which occurs either because the patient has no significant other to designate, or because the significant other fails to respond to the questionnaire. Especially with mailed questionnaires, data loss has been as high as 40-60 percent of the respondents (Ellsworth, 1975).

Using a research team to conduct evaluations, usually by direct patient interviews, is an expensive but highly desirable approach. Trained interviewers can get highly reliable, objective and in-depth information about patient functioning. The high costs of hiring and training research interviewers makes this an infeasible option unless a significant portion of an agency's budget is allocated for evaluation purposes.

Empirical comparisons of the different perspectives offers limited guidance in selecting an informant. Carr and Wittenbaugh (1969) found little agreement among patients, families and therapists about treatment effectiveness. Other research (Ellsworth, 1975) suggests patients' and their significant others' ratings of symptomatic disturbance are positively correlated at a statistically significant level at intake and at follow-up. The correlations between the ratings were higher at follow-up than at intake.

Garfield, Prager and Bergin (1971) systematically examined the relationships among eight different outcomes, including

therapists, client and supervisor global ratings of improvement, and clients, therapists and supervisors pre- and post-treatment ratings of client disturbance. The global ratings of change by the different informants tended to correlate with one another, and to reflect more change than the pre- and post-ratings of client disturbance. The authors recommended as sensitive indexes of change, the client's self-report of depression on the MMPI, and the therapist's global rating of improvement. They also suggested that difference scores may be more objective measures of change than global improvement ratings, as the latter requires a retrospective judgment about change over time which is highly susceptible to measurement error.

In a critique of the Garfield et al. (1971) study, Fiske (1971) and Luborsky (1971) argued that the use of raw change scores may have unduly influenced their results. They suggested adjusting the difference scores to control for the influence of the pre-test score on the difference score. Luborsky (1971) also defended the use of global improvement scores as having strong face validity; the client or therapist may recognize that the client has changed in ways that are quite significant, but this appears as a minor and insignificant change on a questionnaire tapping may different dimensions of behavior. Clearly, the instrument used for the evaluation must be sensitive to patient change; the time period encompassed by the MMPI is probably too broad to pick up the uneven fluctuations in patient functioning.

Each of the authors reviewed in this section came to the same conclusion regarding the selection of an informant--multiple

measures offer more information than measures from a single informant. Given the lack of intercorrelations among measures by different informants, however, the conclusions drawn from the outcome studies must be specific to the outcome criteria employed. Thus, one measure might indicate a positive outcome while others do not. While using multiple outcome measures does not simplify the task of the evaluator, it is a realistic reflection of the complexity of outcome measurement and the developing technology in evaluation research.

Selecting a Follow-up Period

Measuring outcome at a point after completion of treatment is essential to determining whether treatment had a lasting impact. Few studies of outpatient psychotherapy describe how outcome varies according to the timing of the follow-up (Luborsky, 1971; Meltzoff and Kornriech, 1970). Erickson (1975, p. 529) addressed this issue more directly in his review of outcome studies in mental hospitals, when he wrote:

> The shorter the follow-up period, the easier it is to get complete data but the more uncertainty there is as to whether the patient has had an adequate trial period. As the follow-up period grows longer, it becomes harder to determine whether a relapse is due to a deficiency in the hospital treatment program or to unforeseen crises.

The impact of following up inpatients at different points in time is not clear from the research. In general, following up patients from one month to one year after discharge indicates chronic patients are typically restored to a marginal pre-morbid level of functioning (Davis, Diritz and Pasamanick, 1972; Ellsworth, et al., 1968). When

patients have improved, symptom reduction is more apparent than improvement in social functioning (Ellsworth, et al., 1968).

As Schainblatt (1977) noted, there appears to be no rational basis for selecting a follow-up period in designing an evaluation system. Including change measures for both inpatients and outpatients in the present study further complicates the issue, as outpatient treatment usually does not have as definite an ending point as the inpatient's discharge.

The follow-up times selected for the present study were three and six months after treatment was initiated. This allowed for a comparison of adjustment levels of inpatients and outpatients at two points in time after intake. Given a short average length of treatment in the outpatient and inpatient units in the study, most patients should have completed treatment by the first or second follow-up times.

Data Loss

Related to the follow-up issue is the loss of data due to the difficulty of contacting patients at follow-up. Ellsworth (1975) described this problem as one of the most serious threats to the validity of evaluation research, especially with a mailed-questionnaire follow-up. When a significant percentage of patients are not included in the follow-up, the conclusions regarding outcome are limited by the extent to which the "lost subjects" differ from the entire subject population.

Schainblatt (1977) has proposed several procedural techniques to minimize data loss. They include repeated contact of subjects who fail to respond to initial mailed questionnaires, including telephone interviews of subjects. These recommendations have been incorporated in the present study in order to maximize the return rate and minimize the data loss.

Methodology

If patients' conditions are assessed at intake and at a selected follow-up period, the study has incorporated a "one group pretest-posttest design." Campbell and Stanley (1963) concluded that while this design is better than no evaluation at all, it is subject to erroneous conclusions, as it fails to control for threats to internal validity. Any measure of change between the pre- and post-test may be due to history, the occurrence of other change-producing events; or maturation, the biological or psychological processes which systematically vary with time; or testing, the effect of subjects taking a test more than one time.

Another major threat to internal validity which has been discussed extensively in evaluation methodology literature is statistical regression. This is the tendency of persons who have the most deviant scores (high or low) from the mean to regress toward the mean on second measurement. Presumably, the more deviant the score, the larger the error of measurement it contains. Campbell and Stanley (1963) suggested that regression effects are more apparent

for groups selected for their extreme scores than for groups selected for independent reasons which happen to have an extreme mean.

The strongest methodological correction to the "one group pretest-posttest design" is to control for threats to internal validity by randomly assigning patients to a treatment (experimental) and no-treatment (control) group. With random assignment, the effects of maturation, history, testing, and statistical regression are controlled. However, this technique is often unacceptable to administrators of public mental health programs who feel responsible for providing service to all persons requesting it. While true experiments might be possible for evaluations of one-time, innovative programs, they are infeasible for evaluation of ongoing programs.

The non-equivalent control group design has been recommended as an alternative to randomly assigning subjects to an experimental and control group (Campbell and Stanley, 1963; Speer and Tapp, 1976). Although this design does not provide as strong a control for threats to internal validity, it provides a reasonable and feasible alternative to the more stringent random-assignment design. This strategy involves taking pre- and post-test measures on subjects from an available, intact group which may or may not be equivalent to the treatment group on relevant variables, such as socioeconomic status, initial level of disturbance, etc. For example, Speer and Tapp (1976) recommended comparing test scores established for non-patient

population with a patient population to assess the impact of treatment for the patients.[1]

Critics of this design (Campbell and Erlebacher, 1970; Campbell and Boruch, 1975) have argued that using a non-equivalent control group biases the measure of a treatment effect. They contend that different rates of maturation and regression to the mean (presumable "better" rates for the non-equivalent group) will underestimate the gain to the experimental group resulting from the treatment.

Kenny (1975, p. 346) described the problem as follows:

> Given these pre-treatment differences on the dependent variable, interpreting a difference between the experimental and the control group on the post-treatment measure becomes problematic. To interpret a post-treatment difference one must examine the magnitude of the pre-treatment difference, and that difference or perhaps some adjusted difference can be used as a baseline to judge the post-treatment difference. At issue is whether pre-treatment differences should increase, decrease, or remain stable if there is no treatment effect.

In order to measure the treatment effects in the non-equivalent control group design, Kenny recommended one of four statistical analyses: (a) analysis of covariance; (b) analysis of covariance with reliability correction; (c) raw change score analysis; or (d) standardized change score analysis. Alternatively, Nunnally (1975) suggested that using analysis of covariance partials out the effects of pre-test differences or post-test differences. Instead,

---

[1]This type of comparison may be possible with the current data set, as measurement on a nonpatient sample will be taken at two points in time concurrent with pre- and post-treatment measures on the patient sample.

he recommended examining the interaction of the two groups with pre-test and post-test measures.

Measuring change, whether among treated or untreated subjects, is a problematic area which has stirred great controversy among methodologists. Cronbach and Furby (1970) argued against the use of gain scores however they might be adjusted or refined because "such scores are systematically related to any random error of measurement" (p. 68). They point out that if raw difference scores are used, the difference between the "pre" and "post" test scores is related to the "pre-test" score. The portion of change that is a function of the initial level of adjustment must be removed. Different statistical techniques are recommended for this purpose. Meltzoff and Kornreich (1970) have suggested the transformation of initial and final scores into standardized scores. Ellsworth (1975) recommended using residual change as a means of removing the correlation between the pre- and post-test scores with the difference score. Others contend that residualized gain scores use "observed scores" rather than "true scores" in the residualization, and argue in favor of a "true score change" formula for measuring change (Luborsky, 1971; Cronbach and Furby, 1970). For the type of design used in the current study, however, Cronbach and Furby (1970) have recommended using a significance test on the difference between sample means to describe the magnitude of the treatment effect.

Unfortunately, few real life examples comparing the alternative methodological approaches described above appear in the evaluation literature. The controversy surrounding the use of

nonequivalent control groups has been discussed entirely with respect to evaluation of the Head Start Project (Campbell and Erlebacher, 1970), rendering the above arguments applicable only to the evaluation of an educational program for disadvantaged children. The methodological issues may be quite different in evaluating a mental health program; for example, the concept of maturational rate of subjects with respect to learning is dissimilar to the fluctuations in psychological distress of mental patients.

The controversy surrounding the measurement of change also has failed to produce applied studies of how outcomes differ when various statistical techniques are applied to the analysis. Ellsworth (1975) offered a rare demonstration of the high correlations between pre-hospital adjustment and uncorrected outcome and gain scores, and how residual change scores give "base free" measurements of change. He pointed out that if two treatment programs are equally effective but one has initially more disturbed patients assigned to it, gain score analysis will favor that treatment program. If outcome scores are used, the treatment program with better adjusted patients would be favored. Using residual change scores would reflect treatment differences rather than group differences in initial adjustment scores.

Previous Outcome Studies

Studies testing psychotherapy outcomes are reviewed by Luborsky (1971), Meltzoff and Kornreich (1970), and Erickson (1975). These reviews provide information concerning the results that might

be expected from an evaluation of mental health treatment programs. As such, this information should be valuable in structuring and critiquing the design of an evaluation system.

Concerning the effectiveness of treatment programs in general, Meltzoff and Kornreich (1970) concluded:

> Far more often than not, psychotherapy of a wide variety of types and with a broad range of disorders has been demonstrated under controlled conditions to be accompanied by positive changes in adjustment that significantly exceed those that can be accounted for by the passage of time (p. 175) . . . Reviews of the literature that have concluded that psychotherapy has, on the average, no demonstrable effect are based on an incomplete survey of the existing body of research and an insufficiently stringent appraisal of the data . . . On the contrary, controlled research has been notably successful in demonstrating significantly more behavioral change in treated patients than in untreated controls. In general, the better the quality of the research, the more positive the results obtained (p. 177).

Luborsky (1971) was concerned primarily with what factors, including attributes of the patient, therapist, and treatment modality, influence the outcome of psychotherapy. He reviewed 166 studies of the outcomes of outpatient psychotherapy with adult patients undertaken from 1946-1969. In his introduction, he suggested that patients, as a group, will improve in treatment, but that individual patients may or may not improve, depending on certain factors. Identifying those factors is the focus of his article.

The level of pathology prior to treatment was reported in 14 out of 28 studies to be inversely related to outcome. In these 14 studies, the healthier the patient at intake, the greater the change reported at the end of psychotherapy. Of the other 14 studies, 13 reported no significant relationship between initial level of

disturbance and change, and one study indicated that sicker patients changed more in short-term treatment. Meltzoff and Kornreich (1970) reported a similar discrepancy in studies relating initial disturbance level to outcome. An equal number of studies found a positive, negative and no relationship between pre-morbid state and outcome. These results were applicable in studies of schizophrenia and inpatients, as well as psychoneurotics and outpatients. They offered the following explanation of this diversity of findings:

> Severity of maladjustment can be looked at clinically in terms of symptom intensity, duration, pervasiveness or extent of interference with contemporary life; or it can be viewed more broadly as the balance of functional assets and liabilities in an individual's life compared to some estimate of his potential. Some investigators simply define it operationally in terms of psychometric performance. It is easy to see why different research orientations to the meaning of severity can lead to different results and different approaches to prognosis (Meltzoff and Kornreich, 1970, p. 218).

Luborsky (1971) and Meltzoff and Kornreich (1970) reviewed several studies which indicated the presence of strong affect, primarily anxiety and depression, was related to change. Similarly, the number of complaints endorsed on problem checklists was a positive factor. Both of these findings suggest the patient who reports more affect or complaints is asking for help, and thus is probably more ready to change than patients reporting no problems.

Luborsky (1971) reported other patient factors associated with positive change were patient's intelligence, motivation for treatment, younger age and higher educational achievement. Meltzoff and Kornreich (1970) reached a different conclusion regarding the relationship between patient factors and outcome. They pointed out

the lack of adequate controls in these studies and argue that studies demonstrating a positive relationship between patient factors (age, IQ, marital status, education and social class) and outcome of treatment have confounded the variables. They suggested further research, controlling for intervening variables affecting these relationships, should be conducted before definitive conclusions are drawn.

Luborsky's review concluded that treatment factors generally were not strongly associated with outcome, except that combined forms of treatment (e.g. group plus individual therapy, pharmacotherapy plus psychotherapy) were more effective than single forms of therapy. Again, Meltzoff and Kornreich's review concluded that these relationships have not been sufficiently demonstrated. They argued that the methodological weakness in studies comparing individual and group treatment precludes drawing conclusions regarding their relative effectiveness.

Luborsky acknowledged that the outcome criterion being reported affects the results of the study. Most studies he reviewed tended to use only therapist's gross improvement ratings; these studies tended to show greater improvement than those using pre- and post-treatment difference scores on client self-report inventories. While therapist's rating of improvement has been criticized often as a biased measure of change (Garfield, Prager and Bergin, 1971), it is one of the few criterion measures which tends to have consistent significant correlations with other criterion measures (Fiske, et al., 1964; Garfield, et al., 1971).

In his review of outcome studies in mental hospitals, Erickson (1975) highlighted the importance of the type of measure used as the criterion for success. He criticized the use of traditional length-of-stay and return-to-hospital statistics as outcome measures, as they are more reflective of hospital administrative policy than successful or unsuccessful patient adjustment. Similarly, he warned against equating discharge with successful adjustment. The lack of correlation between measures taken in the hospital and post-hospital adjustment (Ellsworth, et al., 1968) is evidence of the fallacy in using discharge rates as an outcome measure. The patient's functioning at discharge is not necessarily predictive of how he will be functioning several months after discharge. This indicates that neither the fact of discharge nor the level of functioning at discharge are satisfactory outcome measures.

Psychosocial measures of adjustment when the patient has returned to the community are recommended as outcome measures of the effectiveness of treatment (Ellsworth, 1975; Erickson, 1975). If made in addition to discharge measures, follow-up measures give information about whether the patient maintained or built on gains made in the hospital.

Erickson's review (1975) concluded the improvement in patient functioning after discharge is usually limited to regaining a marginal, pre-morbid level of functioning. Chronic patients had poorer outcomes than acutely disturbed persons without prior hospitalizations. Some studies found patients who were married and employed steadily prior to hospitalization also had better outcomes.

Inpatient outcome studies indicate improvement in symptomatic disturbance is more likely than improvement in role skills or psychosocial behaviors (Ellsworth, et al., 1968; Pasamanick, et al., 1967). Employment following hospitalization is rarely improved over premorbid functioning, and most discharged patients will not be working full time at follow-up (Anthony, et al., 1972).

Erickson (1975) concluded that most treatment programs, regardless of the type of intervention, have not been able to resotre marginally functioning patients to a normal, community level of psychosocial functioning. Meltzoff and Kornreich (1970) also pointed out that in several experimental studies demonstrating positive outcomes, the gains of the experimental group were either temporary or later matched by advances in the control group. This suggests hospital treatment programs either fail to produce long-term change in patients or the illnesses which repeatedly bring a patient to the hospital are a continuous phenomenon with recurring elevations in symptomatic and behavioral disturbances. While both explanations undoubtedly have merit, the long-term follow-up of patients suggests the latter explanation has more implications for the evaluation of outcomes. If the most successful and innovative inpatient programs have been unable to produce long-term improvement in patient functioning, expecting traditional inpatient programs to effect long-term patient gains is unrealistic.

Drawing generalizations from the outcome literature is complicated by the lack of uniform, intercorrelated outcome measures and the variability in patient characteristics prior to treatment.

Outcomes clearly reflect interactions between the type of hospital treatment program, aftercare services and patient factors such as diagnosis, chronicity and sex. Erickson (1975) recommended future studies become more specific in relating the type of treatment and patient characteristics necessary to achieve the desired treatment objectives within a certain time period.

METHOD

Description of the Study

This study involved testing a series of hypotheses relevant to determining the validity of the BSI. The data were collected as part of a pilot evaluation project sponsored by the Michigan Department of Mental Health, in conjunction with the Clinton-Eaton-Ingham County Community Mental Health Board and St. Lawrence Hospital. The study design originated in an Urban Institute proposal (Schainblatt, 1977), and was subsequently modified by local and state participants in the project. The purpose of the pilot project was to assess the feasibility, cost and utility of a system to monitor the outcomes of state-supported mental health treatment programs.

The pilot evaluation project was conducted in seven local inpatient and outpatient units. The pilot was designed to collect pre- and post-treatment information about a patient's psychological symptoms and social functioning, family burden and patient satisfaction with service. Overlapping information was provided by three different informants: the patient, his or her clinician, and a "significant other" designated by the patient. A non-patient sample was tested with identical self-report instruments and procedures, thus creating a quasi-experimental design with a nonequivalent control group.

Subjects

All adults beginning outpatient or inpatient treatment from August 1977 through October 1977 in several outpatient and inpatient units were eligible to participate in the study. The agencies participating in the study were: Ingham Community Mental Health Center, Clinton County Counseling Center, Eaton County Counseling Center, Mason Mental Health Center, Capitol Area Counseling Center, St. Lawrence Inpatient Unit, and New Riverside Treatment Center. In addition to the outpatient and inpatient samples, a random sample of 825 community residents in the Clinton, Eaton and Ingham counties were asked to complete the BSI in October, 1977 and again in April, 1978. Approximately 80 percent of these residents completed the BSI in 1977; information is not currently available on the second administration in 1978.

The demographic characteristics of the inpatient, outpatient and nonpatient populations in the study are shown in Table 1. All three populations had more women than men participating, with the majority of each group under 40 years of age. Most people had been or were currently married.

The outpatient population had a racial composition similar to the residents (8 percent nonwhite), but the inpatient population had a greater percentage (16 percent) of nonwhites. Outpatients and inpatients generally had less education, poorer employment status and lower incomes than the resident population. The greatest differences among the groups was in their prior use of mental health

Table 1.--Demographic Characteristics of Community Residents and Clients in Tri-County Study.

	Community Residents[1] n = 825	Outpatients n = 615	Inpatients n = 198
Sex:			
Male	37%	34%	43%
Female	63%	66%	57%
Age:			
18 - 29	40%	52%	38%
30 - 39	24%	29%	26%
40 - 49	14%	12%	17%
50 - 59	11%	5%	6%
60 - 69	6%	2%	7%
≥ 70	5%	1%	6%
Marital Status:			
Never Married	24%	26%	30%
Married or Remarried	62%	40%	38%
Widowed, Separated or Divorced	14%	35%	32%
Race:			
White	94%	92%	84%
Nonwhite	7%	8%	16%
Education:			
High School (grades 0-12)	53%	64%	80%
College (grades 13-16)	35%	33%	18%
Graduate (grades 17 and above)	12%	4%	3%
Employment:			
Currently Employed	56%	51%	30%
Unemployed	2%	15%	28%
Not in Labor Force	42%	34%	27%
Other	0%	0%	15%
Income:			
$0 - $3,999	11%	37%	31%
$4,000 - $7,999	15%	15%	18%
$8,000 - $11,999	18%	15%	17%
$12,000 and above	56%	32%	34%
Previous Mental Health Serv:			
Previous Outpatient	5%	41%	18%
Previous Inpatient	1%	11%	47%
Current Outpatient	2%	-	-
None	92%	48%	35%

[1]Includes residents of Clinton, Eaton and Ingham counties who agreed to participate in Community Norm Survey conducted by Department of Mental Health in October, 1977.

services, 6 percent of the residents, compared to 52 percent of the outpatients and 65 percent of the inpatients, had had prior mental health treatment.

Instruments

The symptomatic disturbance of patients was measured by the Brief Symptom Inventory (Derogatis, 1977), the short form of the SCL-90-R, a self-report inventory for symptom patterns of psychiatric and medical patients (Derogatis, Rickels, and Rock, 1976). The major reliability and validation studies have been done on the SCL-90-R, although the correlations on the identical symptom dimensions of the two scales range from .92 to .99 (see Literature Review for more information on the SCL-90).

The BSI consists of 53 items that the respondent is asked whether he/she was bothered by during the past week. Responses are on a five-point scale, ranging from "not at all" to "extremely". The BSI measures psychological symptoms along nine primary dimensions: Somatization, Obsessive-Compulsive, Interpersonal Sensitivity; Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, and Psychoticism. Three global indices of disturbance are calculated: the Global Severity Index (GSI), which combines information on the numbers of symptoms and intensity of perceived distress; Derogatis cited the GSI the best single indicator of the current level or depth of the disorder. The Positive Symptom Distress Index (PSDI) is a pure intensity measure, corrected for the number of symptoms. It functions as a measure of response style, indicating whether the

patient is "augmenting" or "attenuating" symptomatic distress in his style of reporting. The Positive Symptom Total (PST) is the number of symptoms the patient reports as positive (Derogatis, 1977).

The Brief form of the Hopkins Psychiatric Rating Scale (B-HPRS) was used to measure the clinician's perception of the patient's symptomatic disturbance. The long form of the HPRS was designed by Derogatis (1977) to be the "psychiatrist's version" of the SCL-90-R. The first nine symptom dimensions of the HPRS are an analogue to nine dimensions on the BSI and SCL-90-R. The B-HPRS was designed by Derogatis for use in the present study; the changes were some minor alterations in the item-descriptors and an elimination of eight additional dimensions on the HPRS. Thus, the B-HPRS includes the same nine symptom dimensions as the BSI, and one Global Pathology Index scale ranging from "absent" (0) to "extreme" (8) pathology.

The HPRS is designed to be used by psychiatrically sophisticated clinicians (e.g., psychiatrists, psychologists, psychiatric nurses). Each dimension of the scale is defined on the form and represented on a seven-point scale ranging from "none" to "extreme". Three of the seven points on each dimension are further defined by brief clinical descriptors.

Because the HPRS has only recently been developed and is not yet commercially available, formal psychometric properties are not developed. The inter-rater reliability of the B-HPRS was not formally tested; however, two hour training sessions on the instrument were held with all clinicians participating in the study. This allowed

clinicians to discuss the scale and their orientation to completing it, and to rate several sample clients.

Client social functioning was measured on the Personal Adjustment and Role Skills (PARS) questionnaire (Ellsworth, 1975). The PARS is a 61-item questionnaire completed by the patient's designated significant other (usually a family member or friend). Although the same form is administered to males and females, factor analyses indicate the items load differently for men and women. Thus, there are nine primary dimensions for men: Interpersonal Involvement, Depression-Agitation, Anxiety, Confusion, Alcohol-Drug Abuse, Household Management, Relationship to Children, Outside Social, and Employment. For women, the Depression-Agitation and Anxiety dimensions are collapsed into one dimension, Depression-Anxiety, thus creating eight dimensions for women. The respondent is asked whether the patient has, during the past month, been a certain way or performed a certain function. The responses are on a four-point scale, ranging from "rarely" to "always", except on the Social and Employment dimensions, in which the responses describe a certain type of activity, such as "stayed at home this past month" to "often involved in outside activities."

The PARS has been developed over the past ten years, and has been refined in five amended versions. It has been used with hundreds of inpatients and clinic clients in extensive research projects. The psychometric properties of the PARS are well established. The reliability estimates for the PARS test-retest stability range from .66 to .95. The internal consistency (alpha) estimates

ranged from .67 to .94 for hospital patients and .73 to .91 for clinic clients. Validity properties have been established indicating high agreement between ratings by significant others and patients; demonstrations of the instrument's ability to reflect different rates of change for female and male patients and to predict whether the patient is likely to have been hospitalized or seen at a clinic have also been made as part of the validation of this instrument (Ellsworth, 1975).

Data Collection

The data collection was handled by the service agency and the Department of Mental Health (DMH). At the community mental health centers, the secretaries and receptionists were trained to administer the consent forms and BSI questionnaire to incoming clients prior to their first intake session. At the inpatient units, psychiatric nurses and aides administered the consent forms and BSI questionnaires normally within 48 hours of admission. This information was then transmitted to the DMH data coordinator, who logged the data, and sent the PARS questionnaire to the designated significant other. The DMH data coordinator was also responsible for mailing questionnaires to the significant other, clients and former inpatients at the three and six-month follow-up.

The B-HPRS was completed at the time of the patient's intake by the clinician in the outpatient units, and the psychiatric nurses and social workers in the inpatient units.

The study of non-patient residents in Clinton, Eaton and Ingham counties was contracted to a private research firm, Westat, Inc. in Rockville, Maryland. They provided trained interviewers who made telephone contact with 1,000 randomly selected residents. The respondents were asked if they were willing to receive a mailed questionnaire (identical to that completed by the patients) and to provide socio-economic information on the telephone about themselves. Eight hundred and seventy four persons agreed to participate; 75 percent of those people returned the completed questionnaire. A second questionnaire was sent to these residents at a time concurrent with the six-month follow-up of the patient sample.

Most of the procedures for collecting data were developed by the Urban Institute consultants to the project. These procedures were designed to maximize the response rates and minimize the costs of data collection. In order to test the impact of a three and a six-month follow-up on response rates and costs, a random sample of half the outpatients were not followed up at three months. All inpatients were followed up at three months; all outpatient and inpatient subjects were followed up at six months. Table 2 shows the number of subjects who returned questionnaires at each point in the data collection process.

Hypotheses

In examining the BSI's validity in measuring the effectiveness of mental health treatment programs, several hypotheses are proposed as criteria for establishing construct validity, and are

Table 2.--Number of Subjects With BSI, PARS and HPRS Questionnaires.

	Inpatients	Outpatients
Intakes During Study Period	202	617
Subjects with BSI		
Intake	122	518
Three Month Follow Up	49	160
Six Month Follow Up	48	287
Subjects with PARS		
Intake	41	247
Three Month Follow Up	29	84
Six Month Follow Up	18	161
Subjects with HPRS		
Intake	167	326
Discharge	105	N/A

tested. These hypotheses reflect previous research findings on how a valid self-report measure of symptoms should function with a psychiatric and non-psychiatric population.

The first question regarding the BSI's validity is whether the self-report scores differentiate groups of individuals who, on the basis of independent criteria, would be expected to report significantly different levels of disturbance. The three groups participating in this study are differentiated by their inpatient, outpatient or nonpatient status. Each group should report severe, moderate or mile symptomatic disturbance respectively. If the BSI is a valid self-report instrument, its scores should reflect these group differences.

Hypothesis 1: Inpatients will report significantly greater symptomatic disturbance than outpatients and nonpatients at intake.

Hypothesis 2: Outpatients will report significantly greater symptomatic disturbance than nonpatients at intake.

Hypothesis 3: At intake, inpatients will score higher on the psychoticism scales (7-9) than outpatients and nonpatients.

Hypothesis 4: At intake, outpatients will score higher on the neuroticism scales (1-6) than the nonpatients.

The second major area of construct validation of the BSI is its relationship to other measures of symptom disturbance. The Personal Adjustment scales on the PARS (completed by the significant other) and the Hopkins Psychiatric Rating Scale (completed by the clinician) provide independent measures of the patient's symptomatic disturbance.

Hypothesis 5: At intake and follow-up, inpatient self-report of symptomatic disturbance will have a significant, positive correlation with the significant other's report of the patient's symptoms.

Hypothesis 6: At intake and follow-up, outpatient self-report of symptomatic disturbance will have a significant, positive correlation with the significant other's report of the patient's symptoms.

Hypothesis 7: At intake and follow-up, the correlation between the significant other's rating and patient's self-report of symptomatic disturbance will be higher for outpatients than inpatients.

Hypothesis 8: At intake, the self-report of outpatient symptomatology will be positively correlated with the clinician's rating of symptoms.

Hypothesis 9: At intake, the self-report of inpatient symptomatology will be positively correlated with the clinician's rating of symptoms.

Hypothesis 10: At intake, the correlation between the clinician's rating and the patient's self-report of symptomatic disturbance will be higher for outpatients than inpatients.

The last area of investigating the construct validity of the BSI is determining whether the scores differentiate between pre- and post-treatment levels of symptomatic disturbance. The instruments sensitivity to changes in patient's symptoms is an important factor in establishing its validity as a tool for evaluating treatment effectiveness.

Hypothesis 11: Inpatients will report significantly less symptomatic disturbance at follow-up than at intake.

Hypothesis 12: Outpatients will report significantly less symptomatic disturbance at follow-up than at intake.

RESULTS

Subject Participation Rates

The participation rates for subjects completing BSI questionnaires and other informants completing PARS and HPRS questionnaires are given in Table 2 (page 44). Many of the patients who began treatment at the participating outpatient and inpatient centers did not complete questionnaires at intake and follow-up. This data loss raises the question of representativeness of the data collected for the sample of subjects who did participate in the study. Of concern is whether the study gathered data from a biased sample of subjects who were substantially different from patients not participating in the study.

As shown in Table 2, 202 inpatients and 617 outpatients began treatment during the intake period of the study. About 60 percent of the inpatients and 85 percent of the outpatients completed a BSI at intake. Of the patients not completing an intake questionnaire, most simply refused to do so. A smaller portion were unable to complete the BSI because they were too disturbed to do so, or were inadvertently omitted from the study because they were overlooked by the staff handling the intake procedures.

After the intake BSI was completed, subjects were asked to sign a consent form giving permission to be sent a follow-up questionnaire and designating a significant other to be sent a PARS

questionnaire at intake and follow-up. Most of the subjects completing the intake questionnaire agreed to receive a follow-up questionnaire. Only about half of these subjects were willing or able to name a significant other to receive a PARS.

The actual follow-up return rates are also shown in Table 2. For the three-month follow-up, about 160 outpatients and 49 inpatients returned a completed BSI.[1] The six-month follow-up yielded 287 outpatient and 48 inpatient BSI questionnaires. Thus, of the subjects eligible to participate in the study, 24 percent of the inpatients and 47 percent of the outpatients completed intake and six-month follow-up questionnaires. A smaller percentage of inpatients (13 percent) and outpatients (18 percent) had completed BSI questionnaires at intake, three and six months.[2]

Data loss was also significant for the PARS and HPRS (see Table 2, page 44). Many patients were unable or unwilling to designate a significant other, and the PARS questionnaire return rates for significant others was also low. Clinicians completed HPRS questionnaires for about 80 percent of the inpatients and 50 percent of the outpatients. The low percentage of completed HPRS's for outpatients was due to poor staff cooperation at one community mental health center; other centers had about 75 percent completion rates for the HPRS.

---

[1] For reasons outside the control of the author, a random sample of only one half of the outpatients was followed up at three months.

[2] The Urban Institute report on the project (Schainblatt, 1979) gives a more detailed description of response and consent rates.

There are two different samples of subjects in this study--the sample completing intake questionnaires and the sample completing intake and follow-up questionnaires. The extent to which each sample is representative of the entire incoming population of inpatients and outpatients is discussed below.

Representativeness of Intake Sample

The first issue--whether the intake data is generalizable to all outpatients and inpatients beginning treatment at the participating facilities--was examined with a series of chi-square tests on the demographic and diagnostic variables of subjects completing and not completing a BSI at intake (see Table 3). When comparing the demographic characteristics (sex, marital status, employment, education, annual income, age, ethnic status, etc.) of the sample, significant differences emerged on one variable, annual income. Outpatients completing an intake questionnaire had significantly higher incomes than outpatients not completing a questionnaire. For the inpatient sample, a larger percentage of patients completing an intake questionnaire were not married than among the non-participating patients.

Demographic variables are usually considered in determining generalizability because they have an assumed relationship to the mental health of the patient sample. More direct measures of disturbance level collected on the sample include previous mental health service, diagnosis, PPB objective,[1] and the clinician's rating of

[1]This variable is information routinely collected by the Department of Mental Health and reflects the clinician's assessment of the primary treatment objective at intake.

Table 3.--Demographic Characteristics of Subjects Completing and Not Completing an Intake Questionnaire.

CHARACTERISTICS	OUTPATIENTS		INPATIENTS	
	Intake Questionnaire was COMPLETED (n=506)	NOT COMPLETED (n=82)	Intake Questionnaire was COMPLETED (n=110)	NOT COMPLETED (n=72)
Sex:	X^2 = .05		X^2 = .98	
Male	34%	36%	40%	49%
Female	66%	64%	60%	51%
Ethnic Status:	X^2 = 1.34		X^2 = .79	
White	93%	89%	86%	79%
Non-White	7%	12%	14%	21%
Marital Status:	X^2 = 3.05		X^2 = 6.44*	
Married	41%	31%	30%	48%
Not Married	59%	69%	70%	52%
Employment:	X^2 = .21		X^2 = .01	
Employed	51%	54%	35%	36%
Unemployed or Not in Labor Force	49%	47%	65%	64%
Education:	X^2 = .07		X^2 = .01	
High School	64%	63%	80%	80%
College	36%	37%	20%	20%
Previous Mental Health Service:	X^2 = 6.02*		X^2 = 6.36*	
None	50%	37%	29%	48%
Previous Inpatient	9%	15%	52%	39%
Other (outpatient, etc.)	40%	48%	19%	14%
Annual Income:	X^2 = 5.89*		X^2 = 2.30[a]	
Less than $4,000 or Public Assistance	36%	45%	33%	32%
$4,000 - 11,999	30%	34%	38%	29%
$12,000 or more	34%	21%	29%	39%
PPB Objective:	X^2 = 7.82*			
Psychosocial Adjustment	81%	70%		
Crisis Resolution	9%	15%	Not Available	
Rehabilitation/Habilitation	3%	10%		
Maintenance	2%	5%		
Diagnosis:	X^2 = 4.33		X^2 = .39[a]	
Mentally retarded or Organic Brain Syndrome	1%	3%	9%	11%
Psychosis	10%	16%	32%	25%
Neurosis	89%	81%	59%	64%
Age:	X^2 = 5.36		X^2 = 4.98	
18-39	80%	72%	68%	61%
40-59	16%	18%	25%	21%
60 or older	4%	10%	8%	18%

[a] In these categories, missing data resulted in the sample size being reduced to n=55 and n=35 for the Completed and Not Completed groups respectively.

* $p < .05$.

symptomatic disturbance on the HPRS. Statistical tests were conducted on these measures to determine whether the participating subjects were more or less disturbed at intake than the entire population.

For the outpatient sample, a chi-square test revealed significant differences between subjects completing and not completing an intake questionnaire on two of the four initial disturbance level variables--previous mental health service and PPB objective (see Table 3). The differences indicate that the group not completing a questionnaire had more previous mental health treatment and had a treatment objective of crisis resolution, rehabilitation/habilitation or maintenance more frequently than the group completing a questionnaire. The "not completed" group also had a higher percentage (16.0 percent compared to 9.8 percent) of persons with a diagnosis of psychosis than the "completed" group, although this difference was not statistically significant. Unfortunately, too few outpatients not completing a questionnaire were rated on the HPRS to compare the differences on this measure of intake disturbance level (see Table 4).

In summary, the outpatients participating in the study had less previous mental health service, different PPB objectives and lower incomes than outpatients not participating in the study. These findings suggest that the outpatients participating in the study at intake were from a higher socio-economic class and were less disturbed than those outpatients not participating. Feedback from the staffs administering the questionnaires further supports this

Table 4.--HPRS Scores of Subjects Completing and Not Completing A BSI at Intake.

	BSI Completed	BSI Not Completed	t-Test
Outpatients			
Males	3.6 (n=105)	3.6 (n=5)	-
Females	3.4 (n=202)	3.5 (n=13)	_
Inpatients	4.44 (n=116)	4.72 (n=57)	-0.94 (p=.35)
Males	4.04 (n=45)	4.86 (n=28)	-1.78 (p<.05)
Females	4.69 (n=71)	4.59 (n=29)	0.26 (p=.79)

conclusion; they reported often not administering the intake questionnaire to patients who seemed particularly upset or in a crisis state.

The information regarding the representativeness of the inpatient sample at intake is contradictory. Although the staff administering the questionnaires reported that the inpatients participating in the study were the better adjusted, less disturbed portion of the inpatient sample and were not representative of the entire inpatient population, the measures of initial disturbance level are ambiguous.

One measure of previous mental health service revealed statistically significant differences between the inpatients participating and not participating in the study (see Table 3). The participating group had a greater percentage of persons with previous

hospitalizations. This finding is somewhat questionable, however, as there was a high percentage of missing data (about 50 percent) for the treatment objective and diagnosis variables.

The HPRS rating (see Table 4) of inpatients indicates the males not participating were more disturbed than those participating. This difference was statistically significant $t(72) = -1.78$, $p < .05$. Among the female outpatients, however, there were not significant differences in the HPRS ratings.

On the basis of the intake measures of disturbance level of inpatients, it appears that the male inpatients in the study were less disturbed than the male inpatient population in general. The female inpatients in the study do not appear to be a biased sample of inpatients. However, the high percentage of missing data and the statements of staff administering the questionnaires cast some doubt on the validity of this conclusion.

Representativeness of Follow-Up Sample

The second issue is whether the subjects who provided follow-up data were a representative sample of all outpatients and inpatients beginning treatment at the participating units. Chi-square tests were performed on the demographic and intake characteristics of outpatients and inpatients who completed and did not complete follow-up questionnaires (see Tables 5 and 6). Separate analyses were completed for the subjects completing three and six-month questionnaires.

Table 5.--Characteristics of Outpatients Returning and Not Returning Follow-Up Questionnaires.

CHARACTERISTICS	3-MONTH FOLLOW-UP		6-MONTH FOLLOW-UP	
	Returned MHQ	Did Not Return MHQ	Returned MHQ	Did Not Return MHQ
Number of subjects	152*	465*	264*	351*
Sex: Male	26%	37%**	27%	39%**
Female	74%	63%	73%	61%
Annual Income:				
Less than $4,000 or Public Assistance	36%	38%	35%	40%
$4,000 - 11,999	25%	32%	31%	28%
$12,000 or over	40%	30%	34%	31%
Ethnic Status:				
White	93%	91%	92%	92%
Non-White	5%	9%	7%	8%
Marital Status:				
Married	46%	38%	42%	38%
Not Married	54%	62%	58%	62%
Education:				
High School or Less	58%	65%	62%	64%
College or Graduate	42%	35%	37%	35%
Employment:				
Employed	57%	50%	53%	50%
Unemployed or Not in Labor Force	43%	50%	46%	50%
Other	1%	0%	0%	0%
Age: 18 - 39	86%	80%	84%	80%
40 - 59	13%	18%	15%	19%
60 or Older	2%	3%	3%	1%
Previous Mental Health Service:				
None	55%	46%	50%	47%
Previous Inpatient	4%	14%	8%	13%
Other	41%	41%	42%	40%

Table 5.--Continued.

CHARACTERISTICS	3-MONTH FOLLOW-UP Returned MHQ	3-MONTH FOLLOW-UP Did Not Return MHQ	6-MONTH FOLLOW-UP Returned MHQ	6-MONTH FOLLOW-UP Did Not Return MHQ
Clinician Rating of Symptoms at Intake: (a)	(n=93)	(n=232)	(n=152)	(n=173)
None	13%	6%	9%	8%
Mild	42%	47%	43%	47%
Moderate	37%	40%	40%	39%
Severe	9%	7%	9%	7%
Client Report of Symptoms at Intake: (b)	(n=151)	(n=367)	(n=264)	(n=254)
None	20%	18%**	14%	23%**
Mild	49%	41%	45%	41%
Moderate	30%	32%	36%	28%
Severe	1%	9%	5%	9%
PPB Objective:				
Psychosocial	87%	77%**	80%	80%
Crisis Resolution	5%	12%	12%	9%
Rehabilitation	6%	9%	6%	9%
Maintenance	1%	2%	1%	3%

*Sample sizes are lower where indicated.

**Statistically significant differences between categories (X^2, $p < .05$).

(a)Ratings are Global Pathology Index on HPRS (0-1 = none, 2-3 = mild, 4-5 = moderate, 6-8 = severe).

(b)Reports are General Severity Index on BSI (0-.99 = none, 1.00-1.99 = mild, 2.00-2.99 = moderate, 3.00-4.00 = severe).

Table 6.--Characteristics of Inpatients Returning and Not Returning Follow-Up Questionnaires.

CHARACTERISTICS	3-MONTH FOLLOW-UP		6-MONTH FOLLOW-UP	
	Returned MHQ	Did Not Return MHQ	Returned MHQ	Did Not Return MHQ
Number of subjects	44	157*	43	158*
Sex: Male	30%	47%**	40%	44%
Female	71%	53%	61%	56%
Annual Income:	(n=18)	(n=67)	(n=20)	(n=65)
Less than $4,000 or Public Assistance	6%	40%**	20%	37%
$4,000 - 11,999	55%	28%	40%	33%
$12,000 or over	39%	31%	40%	31%
Ethnic Status:				
White	86%	71%**	86%	72%
Non-White	5%	18%	9%	17%
Marital Status:				
Married	35%	37%	34%	38%
Not Married	65%	62%	66%	62%
Education:				
High School or Less	95%	75%**	82%	79%
College or Graduate	6%	25%	18%	22%
Employment:				
Employed	27%	31%	37%	28%
Unemployed or not in Labor Force	59%	55%	56%	56%
Other	14%	15%	7%	17%
Age:				
18-39	62%	64%	75%	61%
40-59	27%	22%	19%	25%
60 or older	12%	14%	7%	15%
Previous Mental Health Service:				
None	27%	37%	28%	36%
Previous Inpatient	49%	48%	46%	49%
Other	24%	16%	26%	15%

Table 6.--Continued.

CHARACTERISTICS	3-MONTH FOLLOW-UP Returned MHQ	3-MONTH FOLLOW-UP Did Not Return MHQ	6-MONTH FOLLOW-UP Returned MHQ	6-MONTH FOLLOW-UP Did Not Return MHQ
Clinician Rating of Symptoms at Intake: (a)	(n=39)	(n=129)	(n=37)	(n=131)
None	5%	4%**	8%	3%**
Mild	5%	23%	8%	22%
Moderate	64%	37%	62%	38%
Severe	26%	36%	22%	37%
Client Report of Symptoms at Intake: (b)	(n=42)	(n=72)	(n=39)	(n=82)
None	7%	18%	10%	16%
Mild	33%	32%	31%	33%
Moderate	31%	30%	33%	29%
Severe	29%	20%	26%	22%

*Sample sizes are lower where indicated.

**Statistically significant differences between categories (χ^2, $p < .05$).

(a)Ratings are Global Pathology Index on HPRS (0-1 = none, 2-3 = mild, 4-5 = moderate, 6-8 = severe).

(b)Reports are General Severity Index on BSI (0-.99 = none, 1.00-1.99 = mild, 2.00-2.99 = moderate, 3.00-4.00 = severe).

For the outpatient follow-up sample (Table 5), significant differences were found between outpatients returning and not returning a questionnaire at three months on Sex, Previous Mental Health Service, Client Report of Symptoms at Intake (the General Severity Index on the BSI) and PPB Objective. These differences indicate the three-month outpatient sample underrepresented males and persons with (a) previous inpatient service, (b) a self-report of severe

symptomatic disturbance and (c) a PPB objective of crisis resolution or rehabilitation. Tests on intake characteristics of the six-month outpatient sample revealed statistically significant differences on only one variable--the client's report of symptoms at intake (the General Severity Index on the BSI). This difference indicates the six-month outpatient sample underrepresented persons reporting no symptoms and severe symptoms and overrepresented those with mild and moderate symptoms.

For the inpatient sample (Table 6), significant differences were found between inpatients returning and not returning a questionnaire at three months on Sex, Annual Income, Ethnic Status, Education and Clinician Rating of Symptoms at Intake (Global Pathology Index on the HPRS). These differences indicate the three-month inpatient sample underrepresented men, persons with incomes below $4,000, non-white persons, and persons with mild and severe symptoms at intake, and overrepresented college-educated persons. For the six-month inpatient sample, only one characteristic--Clinician Rating of Symptoms at Intake--was significantly different for responders and non-responders. This difference indicates the six-month inpatient sample underrepresented persons rated by clinicians as having mild and severe symptoms.

As summarized in Table 7, the sample of outpatients and inpatients who completed <u>both</u> intake and follow-up questionnaires does not appear to be entirely representative of the entire incoming population. Both the outpatient and inpatient samples underrepresent men and persons with severe symptomatology at intake. The

Table 7.--Effects of Data Loss on Tri-County Study.

	Intake	Three-Month	Six-Month
OUTPATIENTS:			
Number of Eligible Subjects	617	617	617
Subjects with completed questionnaires:	506 (82%)[a]	152 (25%)	264 (43%)
Subjects returning questionnaires (when compared with subjects not returning questionnaires) had significantly:*	Higher incomes, less previous mental health service, more "psychosocial adjustment" PPB objectives	Smaller percent of males and persons with previous mental health services, persons reporting of severe symptoms at intake and more "psychosocial adjustment" PPB objectives	Smaller percent of males and persons reporting no symptoms and severe symptoms at intake
INPATIENTS:			
Number of Eligible Subjects	202	202	202
Subjects with completed questionnaires:	110 (55%)[a]	44 (22%)	48 (21%)
Subjects returning questionnaires (when compared with subjects not returning questionnaires) had significantly:*	Smaller percent of married persons; males had lower symptom disturbance ratings by clinicians; more previous mental health service	Smaller percent of males, persons with incomes below $4,000, minorities, and persons rated by clinicians as having mild and severe symptoms at intake, and persons with college education	Smaller percent of persons rated by clinicians as having mild and severe symptoms

[a]Percentages are of the incoming population eligible for the study (617 outpatients, 202 inpatients).

*Unless otherwise noted, significance was determined with chi-square test, $p < .05$.

six-month sample shows less bias on a variety of demographic variables than the three-month sample. For the outpatients, this may have been due to the reduced sample size at three months resulting from a follow-up of only half of the consenting outpatients.

Validity: Differentiation of Groups

The first four hypotheses deal with the extent to which the BSI reflects the differences in symptomatic disturbance expected for self-differentiated groups, i.e., inpatients, outpatients, and non-patients.

Hypothesis 1: Inpatients will report greater symptomatic disturbance than outpatients at intake.

Analysis of variance tests confirmed that inpatients scored significantly higher at intake on all global and dimension scores than outpatients (see Table 8). A main effect was also found for sex, with females scoring significantly higher than males. Significant interaction effects between sex and inpatient/outpatient status were found on three dimension scales (Somatization, Obsessive-Compulsive and Psychoticism) and on two global scales (General Severity Index and Positive Symptom Distress Index).

These results indicate that inpatients do report significantly more symptomatic disturbance at intake than outpatients, but this finding is stronger for females than male inpatients. Table 9 shows the breakdown of BSI intake scores for males and females. A t-test comparing inpatient and outpatient scores by sex reveals that female inpatients have significantly higher scores at intake than

Table 8.--Brief Symptom Inventory Intake Scores for Inpatients and Outpatients.

	Inpatients n=121	Outpatients n=518	F-TEST SCORES FOR MAIN EFFECTS/ INTERACTION		
			Status[a]	Sex	Status/ Sex
BSI Dimensions:					
1. Somatization	1.18	.81	20.9*	11.0*	5.45*
2. Obsessive-Compulsive	1.79	1.37	17.8*	8.3	6.9*
3. Interpersonal Sensitivty	1.71	1.45	6.7*	19.8*	2.3
4. Depression	2.04	1.69	11.5*	17.8*	3.3
5. Anxiety	1.90	1.62	9.0*	22.9*	2.2
6. Hostility	1.44	1.17	8.2*	12.1*	2.1
7. Phobic Anxiety	1.18	.78	22.7*	11.2*	3.1
8. Paranoid Ideation	1.64	1.24	16.3*	7.5*	2.5
9. Psychoticism	1.65	1.23	21.5*	13.2*	6.2*
10. Additional Items	1.92	1.41	25.1*	10.8*	2.9
BSI Global Scores:					
1. General Severity Index	1.63	1.27	23.7*	21.2*	5.8*
2. Positive Symptom Total	32.93	29.46	8.8*	17.7*	0.4
3. Positive Symptom Distress Index	2.42	2.11	23.6*	16.7*	3.7*

*$p < .05$

[a]Status refers to inpatient versus outpatient status.

Table 9.--Brief Symptom Inventory Mean Intake Scores for Inpatients, Outpatients and Community Residents.

	INPATIENTS		OUTPATIENTS		COMMUNITY RESIDENTS	
	Males (n=50)	Females (n=71)	Males (n=174)	Females (n=344)	Males (n=236)	Females (n=417)
BSI DIMENSIONS:						
1. Somatization	.85	1.41*	.71**	.86**	.19	.39
2. Obsessive-Compulsive	1.39	2.07*	1.29**	1.42**	.59	.73
3. Interpersonal Sensitivity	1.31	1.99*	1.23**	1.56**	.48	.65
4. Depression	1.62	2.35*	1.48**	1.79**	.47	.58
5. Anxiety	1.51	2.17*	1.39**	1.73**	.43	.60
6. Hostility	1.13	1.66*	1.02**	1.25**	.45	.49
7. Phobic Anxiety	.89	1.38*	.65**	.84**	.16	.28
8. Paranoid Ideation	1.35	1.84*	1.13**	1.30**	.52	.55
9. Psychoticism	1.26	1.92*	1.11**	1.29**	.30	.36
BSI GLOBAL SCORES:						
1. General Severity Index	1.27	1.89*	1.12**	1.34**	.39	.51
2. Positive Symptom Total	29.6	35.3*	26.7**	30.9**	14.2	16.9
3. Positive Symptom Distress Index	2.16	2.60*	1.99**	2.17**	1.24	1.38

*T-test revealed significantly higher scores ($p < .05$) for inpatients than outpatients.

**T-test revealed significantly higher scores ($p < .05$) for outpatients than community residents.

female outpatients, but that male inpatients do not have significantly higher scores at intake than male outpatients. This finding was true for all global and dimension scores.

Hypothesis 2: Outpatients will report greater symptomatic disturbance at intake than non-patients.

This hypothesis was confirmed. As illustrated in Table 9, outpatients scored higher on each of the BSI global and dimension scores than the nonpatient community sample. These differences were statistically significant (t-test, $p < .05$) for both males and females.

Hypothesis 3: At intake, inpatients will score higher on the psychoticism scales (7-9) than outpatients and nonpatients.

As shown in Table 9, inpatients scored significantly higher than outpatients and nonpatients at intake on the scales reflecting psychotic symptoms (Phobic Anxiety, Paranoid Ideation and Psychoticism). For the Psychoticism scale (#9), there was a significant interaction effect between in/outpatient status and sex. This interaction indicates that the difference between inpatients and outpatients was greater for female than male inpatients.

Hypothesis 4: At intake, outpatients will score higher on the neuroticism scales (1-6) than the nonpatients.

This hypothesis was confirmed for both males and females. Outpatients scored significantly higher on the dimension scales reflecting neurotic symptoms (Somatization, Obsessive-Compulsive, Interpersonal Sensitivity, Depression, Anxiety and Hostility) than the nonpatients. Outpatients also scored higher than the

nonpatients on scales 7-9, the psychoticism scales. These differences were statistically significant (t-test, $p < .05$) for both males and females on all dimension scales.

Validity: Concurrence Between Measures

Hypotheses 5-10 also address the issue of the BSI's construct validity by examining the concurrence between the BSI and two other measures of symptomatic disturbance, the Hopkins Psychiatric Rating Scale (Brief Form) and the symptom scales of the Personal Adjustment and Role Skills questionnaire.

Hypothesis 5: At intake and follow-up, inpatient self-report of symptomatic disturbance will have a significant, positive correlation with the significant other's report of the patient's symptomatic adjustment.

The correlation between inpatient scores on corresponding symptom scales is shown in Table 10. The General Severity Index on the BSI was compared with the Overall Symptomatic Disturbance score on the PARS. Each of the dimension scales on the BSI that has a comparable measure on the PARS was compared with its corresponding scale on the PARS.

At intake, correlations between inpatient scores were generally quite low. None of the correlations were significant at the .05 level. The most important relationship, that between the global measures of symptomatic disturbance on the PARS and BSI (line 1) was .23, indicating a positive but non-significant relationship.

At follow-up, the correlation between measures was higher than at intake. Four of the seven correlations were statistically

Table 10.--Correlations Between Inpatient Scores on Symptomatic Disturbance Scales on the BSI and PARS.

Corresponding Scales		INTAKE	THREE-MONTH FOLLOW-UP	SIX-MONTH FOLLOW-UP
BSI	PARS	Inpatients (n=35)	Inpatients (n=17)	Inpatients (n=8)
1. General Severity Index (G.S.I.)	Overall Symptomatic Disturbance	.23	.58*	.40
2. Anxiety	Anxiety (Males Only)	-.04 (n=12)	.98 (n=3)	.58 (n=3)
3. Psychoticism	Confusion	.02	.20	.70*
4. Interpersonal Sensitivity	Interpersonal Involvement	-.13	.52*	-.28
5. Depression	Agitation-Depression	.26	.58*	.84*
6. Anxiety	Agitation-Depression	.00	.69*	.80*
7. Phobic Anxiety	Agitation-Depression	.24	.32	.67*

*Pearson r, $p < .05$.

significant at the three-month follow-up (n=17). At the six-month follow-up, in spite of a sample size decrease[1] (n=8), four of the seven correlations were significant. The global measures of symptomatic disturbance (line 1) were significantly correlated at three months ($r = .58$, $p < .05$). The same correlation at six months was positive but not significant ($r = .40$, $p = .18$), perhaps due to the small sample size.

Correlations between intake and follow-up scores on the PARS symptom scores were generally low, but significant ($r = .38$, $p < .05$ for outpatients; $r = .36$, $p < .05$ for inpatients). The low correlations indicate significant others reported different levels of symptoms at follow-up than at intake.

In summary, the inpatients' report of symptomatic disturbance was not strongly related to their significant others' report of symptomatic disturbance. The correspondence between reports was greater at follow-up than at intake.

Hypothesis 6: At intake and follow-up, outpatient self-report of symptomatic disturbance will have a significant, positive correlation with the significant others' report of the patients' symptomatic adjustment.

In general, this hypothesis was confirmed at intake and follow-up. As shown in Table 11, the global measures of symptoms on the BSI and PARS had significant, positive correlations of $r = .34$, $r = .29$ and $r = .30$ at intake, three and six months respectively.

---

[1] Sample size decreases resulted from the small number of patients with completed BSI and PARS questionnaires.

Table 11.--Correlations Between Outpatient Scores on Symptomatic Disturbance Scales on the BSI and PARS.

Corresponding Scales		INTAKE	THREE-MONTH FOLLOW-UP	SIX-MONTH FOLLOW-UP
BSI	PARS	Outpatients (n=232)	Outpatients (n=70)	Outpatients (n=119)
1. General Severity Index (G.S.I.)	Overall Symptomatic Disturbance	.34*	.29*	.30*
2. Anxiety	Anxiety (Males Only)	.34* (n=73)	.76* (n=22)	.25* (n=34)
3. Psychoticism	Confusion	.22*	.18	.33*
4. Interpersonal Sensitivity	Interpersonal Involvement	-.04	.17	.05
5. Depression	Agitation-Depression	.43*	.16	.29*
6. Anxiety	Agitation-Depression	.45*	.30*	.32*
7. Phobic Anxiety	Agitation-Depression	.40*	.15	.20*

*Pearson r, $p < .05$.

At intake, six of the seven correlations were statistically significant. At three months, three of the seven were significant, and at six months, five of the seven were significant.

Hypothesis 7: At intake and follow-up, the correlation between the significant other's rating and patient's self-report of symptomatic disturbance will be higher for outpatients than inpatients.

This hypothesis was confirmed at intake, when there was a stronger statistical significance for the correlations between outpatient scores than inpatient scores (see Tables 10 and 11). At intake, almost all of the outpatient correlations were statistically significant, whereas none of the inpatient correlations were significant.

At follow-up, the correspondence between PARS and BSI measures was about equally strong for outpatients and inpatients. The global measures were significantly correlated for both inpatients [$r(17) = .58$, $p < .01$] and outpatients [$r(70) = .29$, $p < .01$] at three months. At the six month follow-up, the global measures of PARS and BSI were: [$r(119) = .30$, $p < .01$] for outpatients, and [$r(8) = .40$, $p = .17$] for inpatients. Although the correlation was higher for inpatients, the difference failed to reach significance, possibly due to the small sample size (n=8) for inpatients.

Hypothesis 8: At intake, the self-report of outpatient symptomatology will be positively correlated with the clinician's rating of symptomatology.

This hypothesis was generally confirmed. The corresponding global measures of symptomatology--the General Severity Index on

the BSI and the Global Pathology Index on the HPRS--were correlated at statistically significant levels for both males [$r(106) = .35$, $p < .05$] and females [$r(215) = .40$, $p < .05$].

As Table 12 illustrates, the dimension scores on the HPRS and BSI were significantly correlated for males on all the dimensions but one, Obsessive-Compulsive, which was correlated at the .07 level. For females, six of the nine dimensions were significantly correlated; the Obsessive-Compulsive, Paranoid Ideation and Psychoticism scales had positive, but not significant, correlations.

Hypothesis 9: At intake, the self-report of inpatient symptomatology will be positively correlated with the clinician's rating of symptomatology.

This hypothesis was generally confirmed for female inpatients, but not for male inpatients. The global measures on the HPRS and BSI were significantly correlated for females [$r(66) = 21$, $p < .05$], but not for males [$r(44) = .16$, $p < .05$].

The dimension scores showed a somewhat erratic pattern of correlations for male inpatients (see Table 12). Three of the nine scales (Somatization, Depression and Anxiety) had significant positive correlations. Hostility had a significant negative correlation [$r(44) = -.33$, $p < .05$]; the other scales had low positive and negative correlations.

For female inpatients, six of the nine scales had significant positive correlations, with Paranoid Ideation, Psychoticism and Phobic Anxiety correlating around the $p = .10$ level.

Table 12.--Mean Scores and Correlations Between Corresponding HPRS and BSI Scales.[a]

Corresponding HPRS and BSI Scales		OUTPATIENTS				INPATIENTS			
		Males (n=106)		Females (n=215)		Males (n=44)		Females (n=66)	
		Mean Score	Correlation Coefficient	Mean Score	Correlation Coefficient	Mean Score	Correlation Coefficient	Mean Score	Correlation Coefficient
1. Somatization	HPRS	1.1	.35*	1.3	.40*	1.1	.29*	1.6	.32*
	BSI	0.7		0.8		0.9		1.4	
2. Obsessive-Compulsive	HPRS	1.1	.14	1.2	.04	1.6	.11	1.4	.26*
	BSI	1.3		1.4		1.4		2.1	
3. Interpersonal Sensitivity	HPRS	2.4	.32*	2.4	.23*	2.6	.16	2.6	.37*
	BSI	1.3		1.5		1.3		2.0	
4. Depression	HPRS	2.1	.43*	2.4	.37*	2.7	.57*	3.1	.37*
	BSI	1.5		1.7		1.7		2.4	
5. Anxiety	HPRS	2.3	.39*	2.5	.28*	2.6	.25*	2.9	.40*
	BSI	1.4		1.7		1.5		2.2	
6. Hostility	HPRS	1.5	.19*	1.5	.19*	1.1	-.33*	1.3	.22*
	BSI	1.0		1.2		1.2		1.7	
7. Phobic Anxiety	HPRS	0.6	.29*	0.5	.25*	0.7	.12	1.0	.12
	BSI	0.7		0.8		0.9		1.4	
8. Paranoid Ideation	HPRS	1.1	.32*	0.6	.05	1.3	.04	1.7	.17
	BSI	1.1		1.2		1.4		1.8	
9. Psychoticism	HPRS	0.9	.18*	0.6	.06	1.3	-.09	1.6	.16
	BSI	1.1		1.2		1.3		1.9	
10. HPRS-Global Pathology Index		3.4	.35*	3.3	.29*	4.0	.16	4.7	.21*
BSI-General Severity Index		1.1		1.3		1.3		1.9	

[a]Scales administered at intake.

*$p < .05$.

Hypothesis 10: At intake, the correlation between the clinician's rating and the patient's self-report of symptomatic disturbance will be higher for outpatients than inpatients.

This hypothesis was confirmed for the male subjects in the study. As described above, the correlations between the two scales were consistently positive and statistically significant for male outpatients. However, male inpatients had erratic positive and negative correlations which generally were not statistically significant.

For female subjects, the correlations between the General Severity Index (BSI) and the Global Pathology Index (HPRS) were slightly higher for outpatients [$r(215) = .29$, $p < .05$] than for inpatients [$r(66) = .21$, $p < .05$]. However, this difference was not statistically significant ($p = .72$). This finding, and that of an equivalent number of dimension scales being significantly correlated for both inpatients and outpatients, suggests that the hypothesis was not confirmed for female subjects.

Thus far, the BSI appears to satisfy several of the validity criteria for outpatients and nonpatients. Outpatients had significantly higher scores than nonpatients on all of the BSI scales. Outpatient reports of symptomatology were significantly correlated with clinicians' ratings of symptomatology at intake. They were also significantly correlated with the significant others' report of clients' symptoms at intake, three- and six-month follow-up.

The BSI meets some, but not all, of the validity tests for inpatients. Inpatients had higher scores on all of the BSI scales

than outpatients, but the difference was statistically significant only for females. The self-report of symptoms was significantly correlated with clinicians' ratings for females but not for males. Inpatient BSI scores were not correlated with significant others' ratings at intake, but had many significant correlations at three- and six-month follow-up. The BSI appears to meet certain validity for female but not male inpatients.

Sensitivity of BSI to Change

Hypotheses 11 and 12 examine the relationship between BSI scores at intake and follow-up. The central question is whether the BSI is sensitive to change in an outpatient and inpatient population between intake, three and six months.

A balanced design Analysis of Variance program[1] was used to examine these hypotheses in order to provide simultaneous tests of three main factors--time, patient status, and sex. The analyses required that all subjects have BSI scores at intake, three and six months. Thus, the sample size is significantly reduced from that in previous hypotheses where only BSI scores at intake were examined. The sample size for the predictive validity hypotheses was 27 inpatients and 113 outpatients.

Hypothesis 11: Inpatients will report greater symptomatic disturbance at intake than at (three or six months) follow-up.

---

[1]BALANOVA User's Manual (Frankmann, 1978).

Hypothesis 12: Outpatients will report greater symptomatic disturbance at intake than at (three or six months) follow-up.

These hypotheses were confirmed for inpatients and outpatients. Significant main effects between scores at intake, three and six months were found for inpatients and outpatients on all global and dimension scores except the PST, Positive Symptom Total (see Table 13). As Table 13 illustrates, inpatients and outpatients had significantly lower BSI scores at three- and six-month follow-up than at intake. The reduction in symptom disturbance was most substantial between intake and three months, although symptoms did continue to decrease between the three and six-month follow-ups.

When male and female scores are analyzed separately for outpatients and inpatients, the women's scores changed more than the mens. Females reported more symptoms at intake than males, and also reported more change at follow-up than males. The different rate of change for females was reflected in the finding of significant interaction effects ($p < .05$) between sex and time on three of the four global scales (the General Severity Index, Grand Total and Positive Symptom Total) and interaction effects approaching significance ($p < .10$) on the dimension scores for Interpersonal Sensitivity, Depression, Anxiety and Psychoticism. These findings indicate that the BSI is more sensitive to change among females than males, regardless of their patient status. Most of the BSI scales reflected a pre- and post-treatment difference in symptoms for females. This finding was not true for males, for whom three of the global scales do not reflect significant reduction in symptoms after treatment.

No significant main effects were found between inpatients and outpatient scores for this sample of inpatients (n=27). This finding is in contrast to the significant difference between inpatient and outpatient scores at intake in testing Hypotheses 1-4. A larger sample size (121 inpatients and 518 outpatients) was used to test Hypotheses 1-4. The larger sample size included more symptomatic inpatients (GSI = 1.63 for 121 inpatients, compared to GSI = 1.49 for this sample of 27 inpatients). As discussed in the section on Data Loss, when the inpatient sample is limited to only those subjects with a BSI at intake, three and six months, some of the more disturbed inpatients are eliminated from the sample. Thus, the sample of inpatients available for this analysis is only a small portion of the original inpatient population; the generalizability of these findings are, therefore, quite limited.

Table 13.--ANOVA of BSI Scores--At Intake, Three and Six Months.

	INPATIENTS			OUTPATIENTS		
	Intake	Three Months	Six Months	Intake	Three Months	Six Months
General Severity Index						
Males	1.17 (n=11)	.91	.71	1.10 (n=27)	.95	.84
Females	1.81 (n=16)	.81	.60	1.22 (n=86)	.84	.74
Total	1.49 (n=27)	.86	.65	1.16 (n=113)	.89	.79
Significant Main and Interaction Effects: Time ($F(2)=14.90$, $p=.00$) Sex/Time ($F(2)=3.00$, $p=.05$)						
Grand Total						
Males	61.0	48.00	37.64	58.19	49.96	44.48
Females	94.31	43.25	31.63	63.86	43.98	39.06
Total	77.66	45.63	34.63	61.02	46.97	41.77
Significant Main and Interaction Effects: Time ($F(2)=14.67$, $p=.00$) Sex/Time ($F(2)=2.94$, $p=.05$)						
Positive Symptom Total						
Males	26.36	26.00	23.33	23.85	27.14	25.56
Females	35.00	24.17	24.44	27.20	22.70	23.64
Total	30.68	25.08	23.89	25.52	24.92	24.60
Significant Main and Interaction Effects: Sex/Time ($F(2)$-2.85, $p=.05$)						
Positive Symptom Distress Index						
Males	1.96	1.65	1.41	1.96	1.75	1.78
Females	2.49	1.82	1.55	2.02	1.68	1.56
Total	2.23	1.73	1.48	1.99	1.72	1.67
Significant Main and Interaction Effects: Time ($F(2)=14.23$, $p=.00$) Sex/Time ($F(1)=4.56$, $p=.03$)						

Table 13.--Continued.

	INPATIENTS			OUTPATIENTS		
	Intake	Three Months	Six Months	Intake	Three Months	Six Months
Somatization						
Males	.61 (n=11)	.53	.42	.56 (n=27)	.53	.36
Females	1.11 (n=16)	.56	.36	.72 (n=86)	.56	.49
Total	.86 (n=27)	.55	.39	.64 (n=113)	.54	.42
Significant Main and Interaction Effects: Time ($F(2)=5.20$, $p=.01$)						
Obsessive-Compulsive						
Males	1.19	.93	.91	1.24	1.14	.90
Females	1.91	.90	.73	1.33	1.00	.89
Total	1.55	.91	.82	1.29	1.07	.89
Significant Main and Interaction Effects: Time($F(2)=7.82$, $p=.00$)						
Interpersonal Sensitivity						
Males	1.20	1.14	.80	1.29	.99	1.09
Females	2.11	.09	.81	1.42	.98	.86
Total	1.66	1.06	.80	1.35	.99	.98
Significant Main and Interaction Effects: Time ($F(2)=8.28$, $p=.00$)						
Depression						
Males	1.61	1.41	.88	1.56	1.28	1.28
Females	2.45	1.20	.80	1.64	1.10	.93
Total	2.03	1.30	.84	1.60	1.19	1.11
Significant Main and Interaction Effects: Time ($F(2)=13.28$, $p=.00$)						
Anxiety						
Males	1.46	1.00	.89	1.36	1.10	.99
Females	2.12	.89	.73	1.65	1.01	.95
Total	1.79	.94	.81	1.50	1.05	.97
Significant Main and Interaction Effects: Time ($F(2)=14.39$, $p=.00$)						

Table 13.--Continued.

	INPATIENTS			OUTPATIENTS		
	Intake	Three Months	Six Months	Intake	Three Months	Six Months
Hostility						
Males	1.30 (n=11)	1.16	.76	.99 (n=27)	.97	.91
Females	1.43 (n=16)	.74	.59	1.15 (n=86)	.78	.71
Total	1.37 (n=27)	.95	.68	1.07 (n=113)	.87	.91
Significant Main and Interaction Effects: Time ($F(2)=6.03$, $p=.00$)						
Phobic Anxiety						
Males	.68	.49	.40	.76	.59	.54
Females	1.36	.55	.38	.68	.41	.44
Total	1.02	.52	.39	.72	.50	.49
Significant Main and Interaction Effects: Time ($F(2)=9.68$, $p=.00$) Sex/Time ($F(1)=4.27$, $p=.04$)						
Paranoid Ideation						
Males	1.22	.75	.76	1.00	1.02	.75
Females	1.71	.69	.63	1.20	.95	.79
Total	1.47	.72	.69	1.10	.98	.77
Significant Main and Interaction Effects: Time ($F(2)=8.76$, $p=.00$)						
Psychoticism						
Males	1.27	.95	.75	1.21	.97	.90
Females	1.85	.76	.38	1.20	.77	.67
Total	1.56	.85	.56	1.21	.87	.79
Significant Main and Interaction Effects: Time ($F(2)=14.37$, $p=.00$)						

Table 13.--Continued.

	INPATIENTS			OUTPATIENTS		
	Intake	Three Months	Six Months	Intake	Three Months	Six Months
			Additional Items			
Males	1.23	.80	.55	1.10	.91	.77
Females	2.19	.95	.63	1.30	.86	.69
Total	1.71	.87	.59	1.20	.89	.73
Significant Main and Interaction Effects: Time ($F(2)=16.19$, $p=.00$)						

DISCUSSION

On the basis of prior research, three criteria were established as the basis for examining the construct validity of the BSI:

1. scores will differentiate between inpatients, outpatients and community residents;
2. scores had concurrence with other measures of symptoms; and,
3. scores were sensitive to pre- and post-treatment differences in symptoms.

The self-report symptom scores collected on a sample of outpatients and inpatients prior to treatment and at follow-up were measured against these criteria. The criteria were largely satisfied for the outpatients, but the findings were more ambiguous for the inpatient sample. These results suggest the BSI has construct validity for an outpatient, but not an inpatient population. This section reviews these findings, discusses alternative theoretical explanations for them, and finally relates them to the larger issue of designing an evaluation system for measuring treatment effectiveness of mental health programs.

Generalizability of Results

In considering the validity tests in this study, it is important to clarify the extent to which these findings are

generalizable to the range of inpatients and outpatients typically seen at public mental health facilities. Clients participating in this study were a sample of all outpatients and inpatients beginning treatment during a three month period at several outpatient and inpatient units. However, the sample of subjects completing BSI questionnaires was not entirely representative of the inpatient and outpatient populations at these centers. As described in the RESULTS section, the subjects participating in this study were less disturbed at intake and had higher socioeconomic characteristics than the agencies' populations. These biases resulted from the more disturbed and lower socioeconomic persons not completing questionnaires at intake and follow-up. Thus, the results of this study are not generalizable to the participating agencies' "typical" outpatient or inpatient populations. Rather, these findings regarding the BSI's validity apply to a somewhat healthier and higher socioeconomic class population than is usually seen at an outpatient or inpatient unit.

Tests of Construct Validity

For the outpatient population, the BSI appears to meet several of the criteria for construct validity. Outpatients scored significantly higher on all BSI scales than the community residents. The outpatients' self-report of symptoms was significantly correlated with the clinicians' rating of symptoms on an analogue scale, the HPRS. In addition, the BSI ratings were significantly

correlated with the significant others' rating of symptoms on the PARS. This finding held up at intake and both follow-up periods.

Although the BSI had strong validity properties for an outpatient population, these findings did not hold up for an inpatient population. For the inpatients, validity of the BSI was established in part for females, but not for male inpatients. When outpatient and inpatient BSI scores were compared, BSI scores were significantly higher for inpatients than outpatients. When sex was controlled for, however, this finding held up for women and not men. Thus, female inpatients reported significantly more symptomatology than female outpatients, but the male inpatients did not report significantly higher symptoms than male outpatients.

A similar sex difference was found in the relationship between inpatients' BSI scores and clinicians' ratings. The scores were significantly correlated for female inpatients but not male inpatients. Thus, the correspondence between clinicians' ratings of symptoms and patients' self-reports was greater for female inpatients than male inpatients. These two findings suggest the BSI is a valid instrument for female inpatients, but not male inpatients.

The correspondence between the inpatients' self-report of symptoms and their significant others' ratings was quite low, especially at intake. The comparisons of global BSI and PARS symptom scales revealed positive but nonsignificant correlations at intake and follow-up. Only the correlation at three-month follow-up achieved statistical significance. The small sample size for this analysis undoubtedly contributes to the low significance level of

these correlations. Unfortunately, the small sample size precluded separate analyses for male and female inpatients. Thus, it is not known whether the lack of validity established earlier for male inpatients was also true in comparing inpatients' BSI and PARS scores.

To determine whether the BSI is sensitive to pre-post treatment change, the difference between scores at intake and follow-up was analyzed for those outpatients and inpatients who completed an intake, three and six-month follow-up questionnaire. When analyzed together, inpatients and outpatients had significantly lower scores at follow-up than intake. When sex and patient status were controlled for, however, the analysis revealed different patterns of change for males and females, outpatients and inpatients.

Although inpatients and outpatients as a group had significantly lower symptomatology at follow-up than at intake, there was a significant interaction effect between sex and time on most of the global scales. Females had a greater pre-post treatment reduction in symptoms than males, regardless of patient status. This finding may have been due to the women having higher intake scores than men, thus allowing for a greater reduction in symptoms at follow-up. These results suggest once again, however, that the BSI is a more valid instrument for females than males.

The inpatient sample[1] used in this analysis did not have significantly higher intake scores than the outpatient sample,

---

[1]Due to the requirement that only subjects with complete data sets (i.e., intake, three and six-month questionnaires completed) be

indicating this inpatient sample was an even less disturbed sample than that used in the first two tests for construct validity. Given this, the finding of insensitivity to pre-post treatment differences is generalizable only to the least disturbed portion of the inpatient population, i.e., those inpatients most closely resembling outpatients on intake measures of symptomatic disturbance.

In summary, the BSI meets most of the established tests for construct validity for the outpatient sample, thus suggesting it may be an adequate instrument for measuring outpatients' symptoms before and after treatment. The BSI does not appear to be sensitive to pre- and post-treatment differences in symptoms for male outpatients, however, and thus may not be valid for this population. If the BSI is used to evaluate outpatient program effectiveness, caution should be used in generalizing any findings to an entire outpatient population. The most disturbed members of the population may be unwilling or unable to complete the BSI, thus limiting the external validity of the evaluation results.

This study indicates the BSI is not a valid instrument for measuring inpatient symptoms before and after treatment. The major problem in using the BSI is the high percentage of data loss, resulting in only 20 percent of the population completing pre- and post-questionnaires. This sample appears to be the less disturbed members of the inpatient population.

---

used for this analysis, the sample size was significantly reduced from that used in the previous analyses. This was especially true for the inpatient sample, which was reduced to 11 males and 16 females.

For those inpatients who do complete the BSI, it appears to be a valid measure of symptoms for females but not males. Males do not report greater symptoms than their outpatient counterparts, nor do they report a reduction in symptoms after treatment. They also report themselves as being less symptomatic than do their clinicians. Thus, the BSI fails all of the tests for construct validity for male inpatients. By contrast, it meets most of the tests for female inpatients. The only exception to this was a lack of correspondence between self and significant others ratings at intake. The small sample size of inpatients available for this analysis limits the strength of these findings, and suggest the need for further testing with a larger sample size.

Alternative Theoretical Explanations

This study found clear differences in men and womens' responses to the BSI. Females consistently report greater symptomatology than males; this occurred among all three groups of subjects--inpatients, outpatients and community residents. The BSI failed to meet any of the criteria for construct validity for male inpatients, and yet, satisfied most of the criteria for female inpatients. This section considers alternative theoretical explanations for these findings.

Interpretation of Sex Differences

Other studies of self-report inventories have also found sex differences, usually in the direction of women reporting more

symptoms than men. Many scales, including the SCL-90 and the MMPI, have incorporated different norms for men and women in their scoring systems (Derogatis, 1977; Welsh and Dahlstrom, 1956). Several studies on different MMPI scales have shown women (including clients and normals) reporting more symptoms than men, particularly anxiety and depression (Welsh and Dahlstrom, 1956). In their extensive review of psychological studies of sex differences, Maccoby and Jacklin (1974) found consistent evidence of greater anxiety among females than males; these studies were conducted on non-patient adults and children.

Maccoby and Jacklin (1974) question whether women are actually more symptomatic than men. They point out that the scales used in these studies were not validated against behavioral or physiological measures of anxiety. They also cite some evidence that males tend to be more defensive than females, and have higher "lie" scale score (Sarason, 1960), and suggest that girls may be conditioned to admit weakness more readily than boys.

Clearly, more research is needed to determine whether the sex differences found in this study and others reflect true differences in the level of symptomatology between men and women. To the extent this study has used independent, objective measures of symptoms (see Discussion section, Concurrence Among Informants), the evidence on this question is mixed. For the outpatient sample, no significant differences occurred between men and womens' symptoms either in the clinicians' or significant others' ratings. And yet, outpatient females rated themselves as more symptomatic than did the male

outpatients. Among inpatients, females rated themselves as more symptomatic than males, as did also the clinicians and significant others.

Several problems arise with this analysis. First, the comparisons are made between ratings on different sets of subjects (see Tables 10, 11 and 12). Secondly, inter-rater reliability of HPRS ratings was not established across clinicians participating in this study. And, finally, the lack of concurrence among different raters and the low absolute values of the correlations between measures must be considered. Definitive evidence that women are truly more symptomatic than men would have to be shown in a study designed more particularly to address this issue than the current study.

If women seeking treatment are more symptomatic than men, the issue becomes whether an instrument such as the BSI should be used to evaluate the effectiveness of treatment programs. Men may be seeking treatment for reasons other than severity of symptoms; indeed, results of this study indicate male inpatients report significantly more social functioning problems (e.g., handling their occupations) than either male outpatients or female inpatients (see Appendix A). Social functioning problems may be the more salient factors for men seeking inpatient treatment, and symptomatic disturbance more salient for women. More research is necessary to address this issue, but current information indicates the BSI alone is not an adequate evaluation instrument, particularly for men.

Measuring Change

Analysis of the change in symptoms from intake to follow-up indicates women change more than men. This was true for outpatient women, as well as inpatient women. The question arises whether this finding was a statistical artifact resulting from the higher intake scores of women. Greater intake scores would allow for more change; this might occur as a regression to the mean, the tendency of deviant scores to move closer to the mean on second measurement.

Campbell and Stanley (1963) described regression to the mean as a serious threat to validity of an evaluation study not using equivalent control groups. They suggested, however, that regression to the mean is more likely to occur when groups are selected for analysis on the basis of their deviant scores.

> Regression effects are thus inevitable accompaniments of imperfect test-retest correlation for groups selected for their extreme scores whenever encountered. If a group selected for independent reasons turns out to have an extreme mean, there is less a priori expectation that the group mean will regress on a second testing, for the random or extraneous sources of variance have been allowed to affect the initial scores in both directions (p. 11, Campbell and Stanley, 1963).

In this study, groups were not selected for analysis because of their extreme scores, and thus, the expectation of a significant regression-to-the-mean factor should be lower. There is substantial evidence, however, that the amount of change reported is directly related to the initial scores. Several studies have reported this and recommend adjusting for the portion of change that is a function of the initial level of adjustment (Ellsworth, 1975; Meltzoff and

Kornreich, 1970). The "best" method for this adjustment is still being debated (Kenny, 1975; Nunnally, 1975; Cronbach and Furby, 1970) and is discussed more thoroughly in the Literature Review.

One frequently recommended technique for controlling for initial scores is analysis of covariance (COVARAN). This study did not use COVARAN to partial out the effects of pretest differences on posttest differences. Nunnally (1975) recommended this technique not be used in quasi-experimental designs because

(a) COVARAN does not control for variables which correlate with both the pretest and posttest variables;

(b) a cardinal assumption regarding equivalent within-groups regression slope is frequently violated; and

(c) there is disagreement about how to adjust statistical parameters of COVARAN to take account of measurement error.

The approach recommended by Nunnally (1975) is to examine the interaction of treatment conditions with comparison groups, using a within subjects, repeated measurement factor. Statistical tests are then performed on all main and interaction effects to determine the significance of the treatment effect and how it interacts with different groups of subjects over time. This approach was utilized in the present study, although other methods of examining change should be undertaken in future studies.

Treatment Effectiveness

Related to the issue of measuring change is whether or not change actually did occur. A plausible alternative explanation for the findings of less change among males than females is that treatment was less effective for men than women.

In their review of psychotherapy research, Meltzoff and Kornreich (1970) concluded

> There is no clear relationship between the sex of the patient and the outcome of psychotherapy. Most studies report no differences. It is typically an incidental finding rather than one upon which research has focused. Much of the data comes from surveys with fallible criteria measures and uncertain experimental conditions (p. 236).

On other measures of change in the present study, there is conflicting evidence whether males change less than females. In responding to a client satisfaction questionnaire at follow-up, males and females had similar responses to questions about how much they had changed as a result of treatment (see Appendix B). Ratings on measures of social functioning indicate a substantial portion of men improved (40 - 50 percent, depending on the informant) in their social functioning after treatment. Further analysis of this issue is needed, using composite measures of change (e.g., analyzing change on symptomatic disturbance and social functioning and client satisfaction measures) to determine if men seek treatment and benefit from treatment in different ways than women.

Concurrence Among Informants

Evidence on the correspondence among different raters (self, clinician and other) indicates greater agreement exists among

measures of symptoms for outpatients than inpatients. Although statistically significant, the correlations were low and once again suggest more correspondence between the BSI and other measures for females than males.

The ambiguous results of this study are not surprising, given that other studies of the relationship between patient, therapist and significant other ratings have had mixed results. Carr and Wittenbaugh (1969) found little agreement among patients, families and therapists about treatment effectiveness. By contrast, two other studies (Ellsworth, 1975; Fontana and Dowds, 1975) found significant correlations between patients' and significant others' ratings at intake and follow-up. In both studies, the correlations were significantly higher at follow-up (range from .50 to .69) than at intake (.20 - .22). It is interesting to note the similarity between these studies and the present in their findings of stronger relationships between ratings at follow-up than at intake. This indicates patients and their significant others have more discrepant views of the patients' symptomatology when the patient enters the hospital than when he has completed treatment.

There were two main differences between the present study and those by Ellsworth and Fontana and Dowds. Their studies had over 200 subjects, thus possibly rendering the low intake correlation statistically significant (an equivalent correlation in the present study was not significant). In addition, their studies used the same scale (PARS) for both the patient and the other. Their use of the same scale, and the possible similar response set this would

engender, might lead to a greater correlation between patient and others' report of symptoms than was found in the present study.

In summary, several theoretical arguments may be raised as alternative explanations to the findings in the present study. Sex differences on the BSI may be due to mens' tendencies to experience and/or report fewer symptoms than women. Findings of greater change among women may occur because the measurement of change has not controlled for womens' higher intake scores. On the other hand, men may change less because the treatment is less effective for them, or because social functioning problems are more salient factors in their treatment than symptoms. Different measures and different informants also raise questions about which are most appropriate for what types of analyses.

Further research is needed to answer definitively the above questions. The evidence in the present study is that the BSI meets tests of construct validity for outpatients but not for inpatients, particularly for male inpatients. The reasons for this may be embedded in the above theoretical positions; the issues are confounded with one another and need further exploration. For whatever reasons the BSI fails certain tests of validity, its use as an evaluation instrument with inpatients must be seriously questioned.

The Evaluation System

The adequacy of a state-wide evaluation system based on a self-report instrument should be assessed separately from the validity of the questionnaire. The data collection procedures and

instruments used in this pilot study together comprise a system for evaluating treatment outcomes. Other procedures and instruments have been pilot tested in other counties in Michigan (Clifford, 1978; Ihlevich, 1979) and will be compared with results of this pilot to develop a possible state-wide strategy for evaluating mental health services in Michigan.

The extent of data loss in this study raises questions about the adequacy of an evaluation system built on a self-report instrument. About 40 percent of the inpatients and 20 percent of the outpatients did not complete intake questionnaires. One report on this project by the Urban Institute (Schainblatt, 1978) suggests the consent rates might be improved by having better trained data collectors in the outpatient units and by having more staff resources committed to collecting data in the inpatient units. Rates might also be improved if agencies had more of an incentive to have clients complete questionnaires. In the current study, many of the staff participating in the study did not believe the information being collected would be useful to them, and cooperated with the study primarily because it was mandated by the Department of Mental Health.

These modifications might result in somewhat higher questionnaire completion rates at intake. The fact remains, however, that large portions of the client populations, especially among inpatients, will be unable or unwilling to complete questionnaires about their psychological and social functioning prior to beginning treatment. The experience in this pilot also indicates many clients are unwilling or unable to complete follow-up questionnaires as well. Only

one-fourth of the inpatients and one-half of the outpatients eligible for this study completed both intake and follow-up questionnaires. The additional problem of lack of instrument validity for those inpatients who do complete the questionnaire argues against the sole use of a self-report model to evaluate mental health services, especially for the inpatient programs.

The self-report model for evaluating mental health programs for outpatients does appear to have validity for the healthier, less disturbed portion of the outpatient population. If care is taken not to generalize the results of the evaluation to the entire outpatient population, the information provided with the self-report model does appear to be useful. In selecting a state-wide evaluation system, however, it is important to be able to compare the effectiveness of different types of programs with the entire range of clients using the services. When the primary objective of public mental health programs is to find cost-effective alternatives to inpatient hospitalization (i.e., outpatient, aftercare and partial hospitalization programs), the evaluation system should allow comparisons across these programs. Because of its limited applicability to more seriously disturbed inpatients and outpatients, the self-report model would not provide for these inter-program comparisons.

An evaluation system based on a clinician's rating of psychological and social functioning constitutes a viable alternative to a self-report model. Having clinicians complete standardized assessment instruments at intake and discharge would insure complete

coverage of both an inpatient and outpatient population. In a separate paper, the clinicians' rating system is recommended as the "preferred alternative" for a state-wide evaluation system (LaFerriere, 1979).

Recommendations for Further Research

In examining the validity of the BSI, several important issues arose concerning the development of a system to evaluate mental health services. In addition, the finding of sex differences in the self-report of symptoms raised several interesting questions for further research. Some of these issues might be explored with the current data set; others would require additional data to be collected.

First, the question around measuring change and treatment effectiveness needs to be explored further. When the "follow-up" data for the community norm sample (i.e., the repeat administration of the BSI at a time concurrent with the six-month follow-up of clients) becomes available for analysis, pre-post treatment differences for clients can be compared with change rates for a normal sample. This quasi-experimental design utilizing a non-equivalent control group will provide a more rigorous test of treatment versus non-treatment effects.

Further analysis is needed to determine whether women are more symptomatic than men, and how these differences are related to seeking and benefitting from mental health services. A factor analysis exploring which of the social functioning and symptom

disturbance variables are most highly associated with male and female treatment outcomes would be very useful.

Finally, the relationship among different informants' ratings of symptoms and social functioning should be re-analyzed using additional data collected during a second pilot phase of this project at the Michigan Institute for Mental Health. The increased sample size from this data set might permit analysis using only those outpatients and inpatients who had self-ratings as well as clinician and significant other ratings.

Conclusions

In pilot testing an evaluation system based on a client self-report instrument, this study found the Brief Symptom Inventory to be a valid instrument for use with female outpatients and inpatients. For these populations, the BSI met three essential criteria for construct validity--differentiation between groups of inpatients, outpatients and nonpatients; congruence with other measures of symptomatology; and sensitivity to pre- and post-treatment differences in symptoms. By contrast, the BSI did not meet the essential validity criteria for males. Male inpatients did not report significantly more symptoms than male outpatients, nor did either sample of men report a lessening of symptoms after treatment. The correspondence between the BSI and reports of symptoms by clinicians and significant others was greater for outpatients than for inpatients, with the lowest correspondence being for male inpatients.

The evidence from this study suggests that while the BSI may be a valid instrument for use with outpatients, it is not valid for use with inpatients. The construct validity of this scale is strongest for female outpatients, and weakest for male inpatients. The combined results of a high percentage of data loss and the lack of instrument validity for males render this self-report model inappropriate for evaluating the effectiveness of mental health services.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abeloff, M.D. and Derogatis, L.R. "Psychological Aspects of the Management of Primary and Metastatic Breast Cancer." Proceedings of the International Conference on Breast Cancer. New York: A.R. List, Inc., 1977.

Anastasi, A. Psychological Testing, 3rd Edition. London: MacMillan Company, Collier-MacMillan Ltd., 1968.

Anthony, W.A., Buell, G.J., Sharrett, S., and Althoff, M.E. "Efficacy of Psychiatric Rehabilitation." Psychological Bulletin 78 (1972): 447-456.

Arnoff, Franklin. "Social Consequences of Policy Towards Mental Illness." Science 188 (1975): 1277.

Boleloucky, L., and Horvath, M. "The SCL-90 Rating Scale: First Experience with the Czech Version in Health Male Scientific Workers." Activitas Nervosa Superior 16 (1974): 115-116.

Campbell, D.T., and Boruch, R.F. "Making the Case for Randomized Assignments to Treatments by Considering the Alternatives: Six Ways in Which Quasi-Experimental Evaluations in Compensatory Education Tend to Underestimate Effects." In Evaluation and Experiment: Some Critical Issues in Assessing Social Programs. Edited by C.A. Bennett and A.A. Lumsdaine. New York: Academic Press, Inc., 1975.

Campbell, D.T., and Erlebacher, S. "How Regression Artifacts in Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful." In Compensatory Education: A National Debate, Vol. 3 of The Disadvantaged Child. Edited by J. Helmuth. New York: Brunner/Mazel, 1970.

Campbell, D.T., and Stanley, J.C. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally, 1963.

Carr, J.E., and Whittenbaugh, J. "Sources of Disagreement in the Perception of Psychotherapy Outcomes." Journal of Clinical Psychology 25 (1969), 1: 16-21.

Clifford, D.L. "Implementing Outcome Evaluation in a Community Mental Health Center." Paper presented at 86th Convention of American Psychological Association. Toronto, 1978.

Craig, T.J., and Abeloff, M. "Psychiatric Symptomatology Among Hospitalized Cancer Patients." American Journal of Psychiatry 131 (1974): 1323-1327.

Cronbach, L.J. Essentials of Psychological Testing. Third Edition. New York: Harper and Row, 1970.

Cronbach, L.J., and Furby, L. "How We Should Measure 'Change' -- Or Should We?" Psychological Bulletin 74 (1970), 1.

Davis, A.E., Dinitz, S., and Pasamanick, B. "The Prevention of Hospitalization Schizophrenia: Five Years After an Experimental Program." American Journal of Orthopsychiatry 42 (1972): 375-388.

Derogatis, L.R. SCL-90 Revised Version Manual-I. John Hopkins University School of Medicine, 1977.

Derogatis, L.R., and Cleary, P.A. "Factorial Invariance Across Gender for the Primary Symptom Dimensions of the SCL-90." British Journal of Social and Clinical Psychology, 1977 (in press).

Derogatis, L.R., Lipman, R.S., and Covi, L. "SCL-90: An Outpatient Psychiatric Rating Scale - Preliminary Report." Psychopharmacology Bulletin 9 (1973),1: 13-17.

Derogatis, L.R., Lipman, R.S., Covi, L., and Rickels, K. "Neurotic Symptom Dimensions as Perceived by Psychiatrists and Patients of Various Social Classes." Archives of General Psychiatry 24 (1971): 454-464.

_____. "Factorial Invariance of Symptom Dimensions in Anxious and Depressive Neuroses." Archives of General Psychiatry 27 (1972): 659-665.

Derogatis, L.R., Lipman, R.S., Rickels, K., Uhlenhuth, E.H., and Covi, L. "The Hopkins Symptom Checklist (HSCL): A Measure of Primary Symptom Dimensions." In Psychological Measurements in Psychopharmacology. Edited by P. Pichot. Basel: Karger, 1974.

Derogatis, L.R., Rickels, K., and Rock, A.F. "The SCL-90 and the MMPI: A Step in the Validation of a New Self-Report Scale." British Journal of Psychiatry 128 (1976): 280-289.

Edwards, D.W., Yarvis, R., Mueller, D.P., Zingale, H., and Wagman, W.J. "Test-Taking and the Stability of Adjustment Scales: Can We Assess Patient Deterioration?" Evaluation Quarterly 2 (May 1978), 2: 275-291.

Ellsworth, R.B. "Consumer Feedback in Measuring the Effectiveness of Mental Health Programs." In Handbook of Evaluation Research. Edited by M. Guttentag and E.L. Stovening. London: Sage Publications, 1975.

_____. PARS Scale Manual. Roanoke: Institute for Program Evaluation, 1975.

_____. "An Evaluation of Mental Health Services." Presented at 4th International Symposium of the Kittay Scientific Foundation, New York, 1976.

Ellsworth, R.B., Foster, L., Childers, B., Arthur, G., and Kroeker, D. "Hospital and Community Adjustment as Perceived by Psychiatric Patients, Their Families and Staff." Journal of Consulting and Clinical Psychology 32 (1968), 5, Part 2, Monograph Supplement.

Erickson, R.C. "Outcome Studies in Mental Hospitals: A Review." Psychological Bulletin 82 (1975), 4: 519-540.

Fiske, D.W. "The Shaky Evidence Is Slowly Put Together." Journal of Consulting and Clinical Psychology 37 (1971), 3: 314-315.

Fiske, D.W., Cartwright, D.S., and Kirtner, W.L. "Are Psychotherapeutic Changes Predictable?" Journal of Abnormal and Social Psychology 69 (1964): 418-426.

Fontana, A.F., and Dowds, B.N. "Assessing Treatment Outcome." Journal of Nervous and Mental Disease 161 (1975), 4: 221-238.

Forsyth, R.P., and Fairweather, G.F. "Psychotherapeutic and Other Hospital Treatment Criteria: The Dilemma." Journal of Abnormal and Social Psychology 62 (1961), 3: 589-604.

Garfield, A.L., Prager, R.A., and Bergin, E. "Evaluation of Outcome in Psychotherapy." Journal of Consulting and Clinical Psychology 37 (1971), 3: 307-313.

Helmstadter, G.C. Principles of Psychological Measurement. New York: Appleton-Century-Crofts, 1964.

Hoffman and Overall. "Factor Structure of the SCL-90 in a Psychiatric Population." Journal of Consulting and Clinical Psychology 46 (1978), 6.

Ihilevich, D. and Glaser, G. Manual for the Progress Evaluation Scales. Owosso, Michigan: Shiawassee County Community Mental Health Services Board, 1979.

Kenny, D.A. "A Quasi-Experimental Approach to Assessing Treatment Effects in the Nonequivalent Control Group Design." Psychological Bulletin 82 (1975), 3: 345-362.

LaFerriere, L. "Recommendations for Developing a Statewide Mental Health Evaluation System." Lansing, Michigan: Department of Mental Health, 1979.

Lick, J. "Statistical vs. Clinical Significance in Research on the Outcome of Psychotherapy." International Journal of Mental Health 2 (1973), 2: 26-37.

Luborsky, L., Chandler, M., Auerbach, A.H., Cohen, J., and Bachrach, H.M. "Factors Influencing the Outcome of Psychotherapy: A Review of Quantitative Research." Psychological Bulletin 75 (1971), 3: 145-185.

Magnusson, David. Test Theory. Massachusetts: Addison-Wesley Co., 1969.

Maguire, L. "Evaluating Treatment Outcome." Ann Arbor, Michigan: Washtenaw County Community Mental Health Center, 1977.

McPhee, C., and Zusman, J. Quality Evaluation of Mental Health Services. Washington, D.C.: National Institute of Mental Health, 1976.

Meltzoff, J., and Kornreich, M. Research in Psychotherapy. New York: Atherton Press, Inc., 1970.

Mikawa, J.K. "Evaluation in Community Mental Health." In Advances in Psychological Assessment, Volume 3. Edited by P. Reynolds. San Francisco: Josey-Bates Publications, 1975.

Nunnally, J.C. Psychometric Theory. New York: McGraw-Hill, 1967.

_____. "The Study of Change in Evaluation Research: Principles Concerning Measurement, Experimental Design, and Analysis." In Handbook of Evaluation Research. Edited by E.L. Stovening and M. Guttentag. London: Sage Publications, 1975.

Pasamanick, B., Scarpetti, F., and Kinitz, S. Schizophrencis in the Community: An Experimental Study in the Prevention of Hospitalization. New York: Appleton-Century-Crofts, 1967.

Paul, G.L. Insight vs. Desensitization to Psychotherapy. California: Stanford University Press, 1966.

Sarason, S.B., Lighthall, F.F., Davidson, K.S., Waite, R.R., and Ruebush, B.K. Anxiety in Elementary School Children. New York: Wiley, 1960.

Schainblatt, A.H. Monitoring the Outcomes of State Mental Health Treatment Programs: Some Initial Suggestions. Washington, D.C.: The Urban Institute, 1977.

_____. "Interim Report: What Happens to the Clients? Monitoring the Outcomes of State and Local Mental Health Services." Washington, D.C.: The Urban Institute, 1978.

Speer, D.C., and Tapp, J.C. "Evaluation of Mental Health Service Effectiveness: A 'Start Up' Model for Established Programs." American Journal of Orthopsychiatry 46 (1976), 2: 217-228.

Waskow, I., and Parloff, M.B. Psychotherapy Change Measures. Rockville, Maryland: National Institute of Mental Health, 1975.

Weissman, M.M., Slobetz, F., Prusoff, B., Mezritz, M., and Howard, P. "Clinical Depression Among Narcotics Addicts Maintained on Methadone in the Community." American Journal of Psychiatry 133 (1976): 1434-1438.

Wider, A. The Cornell Medical Index. New York: Psychological Corporation, 1948.

Wise, T.N., and Fernandez, F. "Psychological Profiles of Candidates Seeking Surgical Correction for Obesity." Psychosomatics, 1977 (in press).

APPENDICES

APPENDIX A

PARS SCORES OF OUTPATIENTS AND INPATIENTS AT INTAKE

PARS Scores of Outpatients and Inpatients

at Intake

Average PARS Scores:	FEMALES Outpatients (n=157)	Inpatients (n=24)
Symptom Score (range = 5-20*)	11.8	15.5
Role Skills (range = 3.8*-15)	11.3	8.6
Alcohol/Drug Abuse (range = 4-16*)	5.4	7.4

*These scores indicate the most dysfunctional score.

Average PARS Scores:	MALES Outpatients (n=81)	Inpatients (n=12)
Symptom Score (range = 4.7-18.7*)	10.8	13.3
Role Skills (range = 3.5*-14)	7.9	6.8
Alcohol/Drug Abuse (range = 5-20*)	7.4	8.5

*These scores indicate the most dysfunctional score.

APPENDIX B

CLIENT SATISFACTION WITH SERVICE

Client Satisfaction with Service

	OUTPATIENTS		INPATIENTS	
	Males	Females	Males	Females
Number of Subjects	41	1.0	12	30
Overall Satisfaction (range: 1-5*)	4.3	4.0	3.7	3.3
Reported Change (range: 1-5*)	4.0	3.8	4.2	3.6

*Most satisfied or greatest change reported.

APPENDIX C

CONSENT FORMS AND INSTRUMENTS

Brief Symptom Inventory

Personal Adjustment and Role Skills Questionnaire

Client Satisfaction Survey Form

DMH-1842-R

#_____

CLIENT CONSENT FORM

MENTAL HEALTH STUDY

We would like to know how the people who use our services are feeling and how they are functioning in their daily lives. We also would like to know if the services provided are helpful in dealing with problems. In order to collect this information, we are asking all clients using our services to participate in a study. This study should help us improve our programs over the next several years. There will be no direct, immediate benefits and no risks to you if you agree to participate.

All information gathered in this study will be kept confidential. Your name will not be connected with the information you provide, nor will this information be shown to anyone outside the evaluation staff in our agency and the Michigan Department of Mental Health. The data collected in this study will be used for statistical purposes in the Department of Mental Health and destroyed within five years of its collection.

The study will involve two parts. You may participate in *one or both parts*:

(1) A questionnaire to be sent to you at one or two later times in the year which you would complete and return to the Department of Mental Health. This questionnaire is similar to the Mental Health questionnaires you just completed; and

(2) A questionnaire to be sent to a person you feel knows you well and could say how you are doing on a daily basis. This person may be a relative (spouse, parent or child if over 14 years) or a close friend. The questionnaire would be sent to this person now and at one or two later times in the year. If you wish to see a copy of this questionnaire, please ask the secretary.

We sincerely hope you will agree to participate in both parts of this study, but there will be no penalty to you if you do not. You will receive the same services whether you agree to participate or not. You may withdraw your consent to participate at any time during the study by notifying any staff member at this agency. If you have any questions about the study, or your participation in it, please ask the secretary or your counselor. Thank you for your cooperation.

* * *

PLEASE CHECK THE BOX OR BOXES BELOW THAT INDICATE YOUR CONSENT OR NON-CONSENT AND THEN SIGN YOUR NAME AT THE BOTTOM OF THE PAGE. THANK YOU.

☐ I agree to participate in the study described above and have had an opportunity to have my questions answered. You may send me a questionnaire at one or two later times in the year at the following address:

Address_____

City_____ Zip Code_____

Telephone No. _____

☐ You may send a questionnaire to this person who knows me well:

Name _____ Is this person here with you today? _____

Address _____

City _____ Zip Code _____

Telephone No. _____

☐ I do not wish to participate in this study because _____

_____

***YOUR SIGNATURE _____ DATE _____

B.S.I.*

INSTRUCTIONS: Below is a list of problems and complaints that people sometimes have. Read each one carefully and decide how much that problem bothered you during the past week, including today. Circle the number under the column heading that best describes how much that problem bothered you. For example, the first problem is "Nervousness or shakiness inside." If you have been bothered by that problem a little bit during the last week, you would circle 2 under the second column. Please do not skip any items. If you change your mind, erase the first mark completely. If you have any questions, please ask the secretary.

During the past week, how much were you bothered by:	Not At All	A Little Bit	Moderately	Quite A Lot	Extremely
1. Nervousness or shakiness inside	0	1	2	3	4
2. Faintness or dizziness	0	1	2	3	4
3. The idea that someone else can control your thoughts	0	1	2	3	4
4. Feeling others are to blame for most of your troubles	0	1	2	3	4
5. Trouble remembering things	0	1	2	3	4
6. Feeling easily annoyed or irritated	0	1	2	3	4
7. Pains in heart or chest	0	1	2	3	4
8. Feeling afraid in open spaces	0	1	2	3	4
9. Thoughts of ending your life	0	1	2	3	4
10. Feeling that most people cannot be trusted	0	1	2	3	4
11. Poor appetite	0	1	2	3	4
12. Suddenly scared for no reason	0	1	2	3	4
13. Temper outbursts you could not control	0	1	2	3	4
14. Feeling lonely even when you are with people	0	1	2	3	4
15. Feeling blocked in getting things done	0	1	2	3	4
16. Feeling lonely	0	1	2	3	4
17. Feeling blue	0	1	2	3	4
18. Feeling no interest in things	0	1	2	3	4
19. Feeling fearful	0	1	2	3	4
20. Your feelings being easily hurt	0	1	2	3	4
21. Feeling that people are unfriendly or dislike you	0	1	2	3	4
22. Feeling inferior to others	0	1	2	3	4
23. Nausea or upset stomach	0	1	2	3	4
24. Feeling that you are watched or talked about by others	0	1	2	3	4
25. Trouble falling asleep	0	1	2	3	4
26. Having to check and double-check what you do	0	1	2	3	4

During the past week, how much were you bothered by:	Not At All	A Little Bit	Moderately	Quite A Lot	Extremely
27. Difficulty making decisions	0	1	2	3	4
28. Feeling afraid to travel on buses, subways, or trains	0	1	2	3	4
29. Trouble getting your breath	0	1	2	3	4
30. Hot or cold spells	0	1	2	3	4
31. Having to avoid certain things, places or activities because they frighten you	0	1	2	3	4
32. Your mind going blank	0	1	2	3	4
33. Numbness or tingling in parts of your body	0	1	2	3	4
34. The idea that you should be punished for your sins	0	1	2	3	4
35. Feeling hopeless about the future	0	1	2	3	4
36. Trouble concentrating	0	1	2	3	4
37. Feeling weak in parts of your body	0	1	2	3	4
38. Feeling tense or keyed up	0	1	2	3	4
39. Thoughts of death or dying	0	1	2	3	4
40. Having urges to beat, injure or harm someone	0	1	2	3	4
41. Having urges to break or smash things	0	1	2	3	4
42. Feeling very self-conscious with others	0	1	2	3	4
43. Feeling uneasy in crowds	0	1	2	3	4
44. Never feeling close to another person	0	1	2	3	4
45. Spells of terror or panic	0	1	2	3	4
46. Getting into frequent arguments	0	1	2	3	4
47. Feeling nervous when you are left alone	0	1	2	3	4
48. Others not giving you proper credit for your achievements	0	1	2	3	4
49. Feeling so restless you couldn't sit still	0	1	2	3	4
50. Feelings of worthlessness	0	1	2	3	4
51. Feeling that people will take advantage of you if you let them	0	1	2	3	4
52. Feelings of guilt	0	1	2	3	4
53. Idea something is wrong with your mind	0	1	2	3	4

DMH FORM NO. 1841R

Agency Use Only:

A.N. |_|_|_|_|

C.C.N. |_|_|_|_|_|_|

N.Q. _____

N.A. _____

Date Comp. |_|_| |_|_| |_|_|

PERSONAL ADJUSTMENT AND ROLES SKILLS QUESTIONNAIRE*

INSTRUCTIONS: Please describe the person's community adjustment during the past month by answering each question below. Mark your answer to each question by circling the number under your answer choice. For example, in question #1, if the person you are rating has shown consideration for you "often" during the last month, you would circle 3 under the heading "often."

Please answer each statement below.

DURING LAST MONTH, HAS HE/SHE . . .	Rarely	Some-times	Often	Always
1. Shown consideration for you.	1	2	3	4
2. Felt close to members of household.	1	2	3	4
3. Discussed important matters with you.	1	2	3	4
4. Been able to talk it through when angry.	1	2	3	4
5. Cooperated (gone along) when things asked of him or her.	1	2	3	4
6. Shown interest in what you say.	1	2	3	4
7. Shown affection toward you.	1	2	3	4
8. Gotten along with other family members.	1	2	3	4

DURING LAST MONTH, HAS HE/SHE . . .	Never	Rarely	Some-times	Often
9. Said people don't care about him/her.	1	2	3	4
10. Said people treat him/her unfairly.	1	2	3	4
11. Complained or worried about problems.	1	2	3	4
12. Said people try to push him/her around.	1	2	3	4
13. Said life wasn't worth living.	1	2	3	4
14. Said things looked discouraging or hopeless.	1	2	3	4
15. Talked about being afraid.	1	2	3	4

DURING LAST MONTH, HAS HE/SHE . . .	Almost Never	Some-times	Often	Almost Always
16. Had difficulty eating (poor appetite, indigestion, etc.)	1	2	3	4
17. Been nervous.	1	2	3	4
18. Acted restless and tense.	1	2	3	4
19. Had difficulty sleeping.	1	2	3	4

DURING LAST MONTH, HAS HE/SHE . . .	Never	Rarely	Some-times	Often
20. Jumped from one subject to another when talking.	1	2	3	4
21. Just sat and stared.	1	2	3	4
22. Forgotten to do important things.	1	2	3	4
23. Been in a daze or confused.	1	2	3	4
24. Needed supervision or guidance.	1	2	3	4
25. Lost track of time.	1	2	3	4
26. Seemed to be off in another world.	1	2	3	4

DURING LAST MONTH, HAS HE/SHE . . .	Never	Rarely	Some-times	Often
27. Been drinking alcohol to excess.	1	2	3	4
28. Been using drugs excessively.	1	2	3	4
29. Become drunk on alcohol or high on drugs.	1	2	3	4
30. Had a drinking or drug problem that upset family.	1	2	3	4
31. Had a drinking or drug problem that interfered with working.	1	2	3	4

DMH FORM NO. 1841R

DURING LAST MONTH, HAS HE/SHE . . .

	Almost Never	Some-times	Often	Almost Always
32. Done chores around house.	1	2	3	4
33. Done household cleaning.	1	2	3	4
34. Prepared meals for the family.	1	2	3	4
35. Done laundry, ironing or mending.	1	2	3	4
36. Done grocery shopping.	1	2	3	4

37. Are there usually children in the home? (Mark one)
(1) _____ No (If "no," skip to question 44)
(2) _____ Yes (If "yes," answer questions 38-43)

DURING LAST MONTH, HAS HE/SHE . . .

	Almost Never	Some-times	Often	Almost Always
38. Spent time with the children.	1	2	3	4
39. Shown affection toward the children.	1	2	3	4
40. Kept promises to the children.	1	2	3	4
41. Been consistent in reacting to the children.	1	2	3	4
42. Known right thing to do when disciplining children.	1	2	3	4
43. Had children show respect for him/her.	1	2	3	4

DURING LAST MONTH, HAS HE/SHE . . .

44. Been involved in activities outside the home? (Mark one)
(1) _____ Stayed at home this past month.
(2) _____ Rarely involved outside the home.
(3) _____ Involved in some outside activities.
(4) _____ Often involved in outside activities.

DURING LAST MONTH, HAS HE/SHE . . .

45. Attended meetings of civic, church or other organizations? (Mark one)
(1) _____ Did not attend any meeting this past month.
(2) _____ Rarely attended meetings.
(3) _____ Sometimes attended meetings.
(4) _____ Often attended meetings.

46. Participated in recreational activities outside the home? (Mark one)
(1) _____ No recreational activities outside home.
(2) _____ Rarely participated in outside recreation.
(3) _____ Sometimes participated.
(4) _____ Often participated.

47. Been employed outside the home? (Mark one)
(1) _____ Unemployed last month (skip to question 51)
(2) _____ Employed part time last month.
(3) _____ Employed full time last month.
Note: If employed part or full time, please answer questions 48-50.

48. About how much take home pay did he/she earn fr working last month? (Do not include money from pension or welfare)
(1) _____ Earned little or no money last month.
(2) _____ Earned less than $100 per week.
(3) _____ Between $100 and $200 per week.
(4) _____ Over $200 per week from working.

49. From working, did he/she earn an adequate amount of money last month?
(1) _____ Earned no money by working last month.
(2) _____ Earned enough to take care of personal needs.
(3) _____ Earned enough to partially support a family.
(4) _____ Earned enough to adequately support a family.

50. Did he/she look forward to going to work each day? (Mark one)
(1) _____ Not employed last month.
(2) _____ Rarely looked forward to work.
(3) _____ Sometimes looked forward to work.
(4) _____ Usually looked forward to work.

DMH FORM NO. 1941R

NOTE: QUESTIONS 51-55 ASK THAT YOU INDICATE WHETHER CERTAIN AREAS OF ADJUSTMENT CAUSED PROBLEMS FOR THE PERSON YOU ARE RATING DURING THE PAST MONTH. PLEASE BE SURE TO ANSWER EACH QUESTION BELOW.

DURING THE PAST MONTH, HAS HE/SHE HAD PROBLEMS . . .

51. a. Talking and relating to you and people close with him or her?
(1) ___ No problems.
(2) ___ Some problems.
(3) ___ Serious problems.
b. If this is a problem, is it . . .
Getting worse ___ Getting better ___ No change ___

52. a. Feeling bad about self or being angry with others?
(1) ___ No problems.
(2) ___ Some problems.
(3) ___ Serious problems.
b. If this is a problem, is it . . .
Getting worse ___ Getting better ___ No change ___

53. a. Being nervous, not sleeping or eating well?
(1) ___ No problems.
(2) ___ Some problems.
(3) ___ Serious problems.
b. If this is a problem, is it . . .
Getting worse ___ Getting better ___ No change ___

54. a. Forgetting things, being confused?
(1) ___ No problems.
(2) ___ Some problems.
(3) ___ Serious problems
b. If this is a problem, is it . . .
Getting worse ___ Getting better ___ No change ___

55. a. Using alcohol or drugs to excess?
(1) ___ No problems.
(2) ___ Some problems.
(3) ___ Serious problems.
b. If this is a problem, is it . . .
Getting worse ___ Getting better ___ No change ___

56. a. Doing household chores, laundry, cooking, cleaning, shopping?
(0) ___ Not expected.
(1) ___ No problem.
(2) ___ Some problems.
(3) ___ Serious problems.
b. If this is a problem, is it . . .
Getting worse ___ Getting better ___ No change ___

57. a. Relating to children in the home?
(0) ___ No children home.
(1) ___ No problem.
(2) ___ Some problems.
(3) ___ Serious problems.
b. If this is a problem, is it . . .
Getting worse ___ Getting better ___ No change ___

58. a. Getting involved in outside social activities?
(0) ___ Not expected.
(1) ___ No problem.
(2) ___ Some problems.
(3) ___ Serious problems.
b. If this is a problem, is it . . .
Getting worse ___ Getting better ___ No change ___

59. a. Earning money from working?
(0) ___ Not expected.
(1) ___ No problem.
(2) ___ Some problems.
(3) ___ Serious problems.
b. If this is a problem, is it . . .
Getting worse ___ Getting better ___ No change ___

HAS THE PERSON YOU ARE RATING RAISED THE FOLLOWING PROBLEMS FOR THE FAMILY DURING THE LAST MONTH? (Circle the number under your answer)

	Not at all	A little bit	Quite a bit	A great deal
60. Has she/he been irritable or angry?	1	2	3	4
61. Have you been worried he/she might hurt themself?	1	2	3	4
62. Have his/her problems CAUSED a drain on the family's finances?	1	2	3	4
63. Have any children in the family been upset BECAUSE of his/her problems (e.g., angry, frightened, sad)?	1	2	3	4
64. Have any children in the family not gotten enough attention BECAUSE OF his/her problems?	1	2	3	4
65. Have you or any other adult in the family been more upset than usual BECAUSE of his/her problems?	1	2	3	4
66. Have you or anyone else in the family taken over extra duties BECAUSE of his/her problems?	1	2	3	4
67. Have his/her problems INTERFERED with your family's activities?	1	2	3	4

PLEASE COMPLETE THE FOLLOWING BACKGROUND QUESTIONS:

68. How often did you see this person during the last month?
(1) ___ Not at all.
(2) ___ Once or twice during past month.
(3) ___ About once a week.
(4) ___ About 3 to 5 times a week.
(5) ___ Saw daily.
(6) ___ Saw daily for 5 or more hours daily.

69. What is your relationship to the person you are rating?
(1) ___ Spouse or mate.
(2) ___ Parent.
(3) ___ Other relative (sister, aunt, etc.)
(4) ___ Friend.

Today's Date _____

(OVER)

6 F/U

CLIENT SATISFACTION QUESTIONNAIRE

The following questions ask you to rate the services you received at our agency. Please circle those answers which best describe your experience at our agency. Your counselor will not see the answers you give here. We will add up the answers all clients give us to help us plan for future services.

1. Did your counselor seem to understand you and your problems?

1	2	3	4	5
Never understood	Seldom understood	Sometimes understood	Usually understood	Always understood

2. Were you satisfied with how often you saw your counselor to get help with your problem?

1	2	3	4	5
Very dissatisfied	Somewhat dissatisfied	Undecided	Somewhat satisfied	Very satisfied

3A. You came to our agency to get help with certain problems. Has there been any change in those problems since you first came to our agency?

1	2	3	4	5
A great deal worse	Worse	Same	Improved	A great deal improved

3B. How important were the services you received at our agency in helping you with those problems?

1	2	3	4	5
Not important at all	Not very important	Undecided	Somewhat important	Very important

On the next few questions, please place a check (✓) beside those answers which best describe your experience at our agency.

4. For each of the following, please check whether you received the service at our agency and how helpful the service was if you received it:

(Check all that apply)	Not Received At this Agency	Not at all Helpful	Not too Helpful	Undecided	Helpful	Very Helpful
a. Group therapy	1. ___	2. ___	3. ___	4. ___	5. ___	6. ___
b. Individual therapy	1. ___	2. ___	3. ___	4. ___	5. ___	6. ___
c. Family therapy	1. ___	2. ___	3. ___	4. ___	5. ___	6. ___
d. Medication	1. ___	2. ___	3. ___	4. ___	5. ___	6. ___
e. Emergency services	1. ___	2. ___	3. ___	4. ___	5. ___	6. ___
f. Marital therapy	1. ___	2. ___	3. ___	4. ___	5. ___	6. ___
g. Activity therapy	1. ___	2. ___	3. ___	4. ___	5. ___	6. ___
h. Respite center services	1. ___	2. ___	3. ___	4. ___	5. ___	6. ___
i. Residential services	1. ___	2. ___	3. ___	4. ___	5. ___	6. ___

(PLEASE TURN OVER TO PAGE 2)

5A. Were you referred to another agency by our staff?

____ 1. YES. The agency recommended to me was _____

_____

____ 2. NO.

5B. If "YES," did you receive services at the agency referred to you?

____ 1. YES ____ 2. NO, because: (Please check all that apply)

____ a. The agency is too far away.
____ b. My problem is solved.
____ c. I wasn't eligible for services there.
____ d. I had to wait too long for an appointment.
____ e. I haven't gotten around to going there yet.
____ f. The fees are too expensive.
____ g. Other (Please explain) _____

_____

6A. Are you still being seen at our agency? ____ 1. YES ____ 2. NO

6B. If "NO," what were the reasons you stopped coming to our agency? (Please check all that apply)

____ 01. My counselor suggested I stop. (Reason: _____)
____ 02. I felt I should stop, but my counselor didn't.
____ 03. My counselor and I both felt that goals were met.
____ 04. My counselor left.
____ 05. No counselors were available at the time(s) I needed to be seen.
____ 06. I felt my therapist was unable to help me.
____ 07. I was sent to another agency for services.
____ 08. The fees were too expensive.
____ 09. I got tired of coming.
____ 10. Other reasons for stopping. (Please explain) _____

_____

Please circle the answers below which best describe your feelings about our agency, as you did in the first section.

7. Would you use our services again if you felt the need?

1	2	3	4	5
No, definitely not	No, I don't think so	Undecided	Yes, I think so	Yes, definitely

8. In general, how satisfied or dissatisfied were you with the services you received at our agency?

1	2	3	4	5
Very dissatisfied	Somewhat dissatisfied	Undecided	Somewhat satisfied	Very satisfied

Please use the space below to share with us any other suggestions, comments or observations that might help us plan our services in the future. Thank you for your cooperation.