

STANDING ON THE SHOULDERS OF A GIANT: DEVELOPMENT OF AN OUTCOME MANAGEMENT SYSTEM BASED ON THE DOSE MODEL AND PHASE MODEL OF PSYCHOTHERAPY

Grant R. Grissom

PsyberMetrics, Inc.

John S. Lyons

Mental Health Services and Policy Program, Northwestern University

Medical School

Wolfgang Lutz

Department of Psychology, University of Berne, Switzerland

Ken Howard's major contributions to psychotherapy research include his studies of the sequence (phase model), rate (dose–response model), and predictability (expected treatment response modeling) of changes that occur during psychotherapy. This body of work formed the foundation for the development of an outcomes management system: Treatment Evaluation and Management (TEaM). The authors describe the development of TEaM measures corresponding to the central constructs of the phase model: subjective well-being, symptoms, and functional disability. Scale scores are combined to derive a composite score, the Behavioral Health Status (BHS) index, which provides a global measure of a patient's psychological status and functioning. The authors present psychometric information about each of the measures and the composite BHS index. Ken was working on finalizing the TEaM at the time of his death. Thus, this measurement approach is his final contribution to the field of outcomes management.

Before his death, Ken Howard was continuing to build on his decades of research on psychotherapy process and outcome to establish a measurement foundation for evaluating the expected trajectory of recovery during treatment. This work was seminal to the emerging field of patient-focused research (Lambert, 2001).

Among his many contributions to the field, studies on the dose–response relation in psychotherapy (Howard, Kopta, Krause, & Orlinsky, 1986) and the phase model of psychotherapeutic change (Howard, Lueger, Maling, & Martinovich, 1993) stand out. The phase model indicates the nature and sequence of changes that occur during treatment. Dose–response research relates the amount of change (“response”) to

The research reported in this article was supported in part by National Institutes of Mental Health Grant 1 R43 MH60029-01. The authors are pleased to acknowledge the assistance of Bruce Briscoe and Sherrill Lord in preparation of this article.

Correspondence concerning this article should be addressed to Grant R. Grissom, PsyberMetrics, Inc., 825 Town Center Drive, Suite 110, Langhorne, PA 19047. E-mail: grant@psybermetrics.com.

the amount of treatment (“dose”). The phase and dose–response research provided the foundation for the first outcomes management system, COMPASS, developed by Ken and his colleagues at Northwestern University in collaboration with Grant R. Grissom and colleagues at Integra, Inc. (Sperry, Brill, Howard, & Grissom, 1996).

The COMPASS system provided longitudinal data necessary to the development of expected treatment response (ETR) models. Ken’s research team showed that it is possible to predict the response of an individual to therapy using measures that correspond to the three phases of treatment response and measures of patient characteristics that mediate treatment response (Howard, Moras, Brill, Martinovich, & Lutz, 1996; Lutz, Martinovich, & Howard, 1999).

Construction of ETR models has direct application both to clinical practice and psychotherapy research (Lueger et al., 2001). Clinicians could compare an individual’s actual progress with the expected course (ETR) to help determine whether treatment is “working.” Research applications are numerous, including evaluation of manualized treatments and improved understanding of the factors determining treatment response/nonresponse.

It is within this conceptual framework that the Treatment Evaluation and Management (TEaM) instrument was developed.¹ We present the conceptual framework, design, and development of the TEaM instrument. In the months before his death, Ken collaborated with the authors of this article to complete the last of three studies necessary to establish the structure, content, and psychometric properties (internal consistency reliability and concurrent validity) of TEaM. The first of these evaluated an initial set of items constructed to reflect the three components of the phase model and items selected as severity adjusters (useful to the construction of ETR curves) or for their potential clinical utility. Provisional scales and subscales were constructed based on the results of this study, and focus groups with practicing clinicians were established to discuss the clinical utility of candidate items. The second study assessed the internal consistency reliability and concurrent validity of the TEaM scales. The final study involved assembly of the full TEaM instrument, construction of full and short-form versions of a global Behavioral Health Status (BHS) scale, determination of the reliability and validity of the BHS, and development of two data quality indicators.

Conceptual Framework of TEaM

On the basis of a meta-analysis, Howard, Kopta, Krause, and Orlinsky (1986) described a dosage model of psychotherapeutic effectiveness that demonstrated a positive relationship between the log of the number of sessions (dose) and the normalized probability of patient improvement (effect). Subsequent dose–response work has provided evidence for the differential, but lawful, responsiveness to psychotherapy of various symptoms (Barkham et al., 1996; Kopta, Howard, Lowry, & Beutler, 1994), interpersonal problems (e.g., Horowitz, Rosenberg, Baer, Ureño, & Villaseñor, 1988; Maling, Gurtman, & Howard, 1995), and diagnoses (Howard, Kopta, Krause, &

¹It was Ken’s wish that the (proprietary) instrument be made available at no cost for noncommercial use by psychotherapy researchers. The authors made that commitment to Ken and will be pleased to honor it. Interested researchers are invited to contact Grant R. Grissom, who will provide updates on the status of TEaM development and documentation for the TEaM system as it becomes available.

Orlinsky, 1986; Pilkonis & Frank, 1988). Basically, the dosage model describes a pattern of relatively rapid early improvement, with more and more sessions needed to achieve incremental improvement later in treatment (a pattern of diminishing returns).

The phase model (Howard, Lueger, Maling, & Martinovich, 1993) extended and interpreted the dosage model, proposing three progressive sequential phases of the psychotherapeutic recovery process: (a) remoralization, the enhancement of well-being; (b) remediation, the achievement of symptomatic relief; and (c) rehabilitation, the reduction of troublesome, maladaptive behaviors that interfere with life functioning (e.g., functioning in areas such as family relationships and work). The phase model posits that the decelerating curve of improvement for a patient can be attributed to the increasing difficulty of treatment goals as they change over the course of treatment. Both the dosage and phase models rely on group data to provide outcome information for an average patient. However, research has shown that patterns of improvement for individuals vary around this general trend (e.g., Barkham, Stiles, & Shapiro, 1993; Krause, Howard, & Lutz, 1998; Martinovich, 1998).

The ETR model assumes an underlying log-linear course of recovery in treatment for each patient, as described in the dosage model. The ETR model uses a hierarchical linear modeling strategy (Bryk & Raudenbush, 1992) to depict a patient's behavioral health status over treatment as a log-linear function of session number; then it uses pretreatment clinical characteristics (e.g., severity, chronicity, previous treatment, treatment expectation) to predict the patient's expected response over the course of his or her treatment (Howard et al., 1996). Using the results of such an individualized growth curve analysis for a large sample of outpatients in psychotherapy, a single patient's course of treatment (ETR) can be predicted as soon as his or her intake information is available. The ETR can be revised based on the patient's initial response to treatment (Lutz, Martinovich, & Howard, 1999). Thus, ongoing therapeutic effectiveness can be assessed for a single patient by tracking the patient's actual progress in comparison to his or her expected progress.

Design Considerations

Predicting and monitoring the trajectory of recovery, with its focus on the clinical management of an individual patient, require new kinds of measures and assessment systems. The instruments must be consumer friendly (e.g., relevant to the patient, not too long, not too complicated) and easy to use in daily practice (several times over the course of treatment), and they must have enough information to support clinical decision making in an ongoing treatment (Sangslund, 2001; Sperry, Brill, Howard, & Grissom, 1996).

The TEaM instrument described in this article is intended for use as a stand-alone measure and as the central component of an outcomes management system for clinical practice. The outcomes management system will include, in addition to the TEaM instrument, software allowing for the efficient collection, storage, processing, and reporting of patient data through multiple administrations concurrently with treatment.

TEaM measurement domains correspond to the three phases of the therapeutic progress established through the phase model: subjective well-being (remoralization, measured by the Subjective Well-Being Scale), symptoms (remediation, measured by the Symptom Checklist), and functional disability (rehabilitation, measured by the Functional Disability Scale). Both the Symptom Checklist and Functional Disability Scale are further divided into subscales. Seven Symptom Checklist subscales corre-

spond to disorders commonly treated in outpatient settings, and three Functional Disability subscales are based on the Social Security Disability Guidelines. In addition, TEaM includes scales for consumer satisfaction with treatment and the therapeutic bond, a malingering index, and a substance abuse screener.

The design criteria for the TEaM outcomes management system reflect the needs of the various stakeholders: patients, clinicians, clinical managers/administrators, case managers/utilization reviewers, and payers. To accommodate patients who may have limited English literacy, the TEaM is designed to be accessible to those who read at the sixth-grade level. Response burden is modest: Most patients will be able to complete the initial administration in less than 15 min, and an abbreviated form requiring less than 5 min is available for treatment monitoring. Patients perceive the TEaM questions as appropriate (face valid). When properly oriented by staff, they consider the assessment to be evidence of the thoroughness and quality of the treatment program.

TEaM is designed for clinical use throughout treatment. Software will be developed to provide for computer administration, followed by immediate printing of a report. The report can be used for treatment planning and monitoring. It can be discussed with the patient during the therapy session to engage the patient, affirm progress, and identify problem areas. The psychometric strength and relevance of the scales enable the clinician to use the reports with confidence. Data quality indicators, including inconsistency checks and a malingering index, help to detect inauthentic responding.

The automated TEaM database is designed to protect patient privacy. Patient names and other identifiers are stored in a file separate from the response data file, allowing for downloading of the response file (e.g., for research or for the development of aggregate reports) without risk to confidentiality. The software addresses other administrative concerns: For example, it provides for easy integration of the TEaM system into office and clinical procedures with negligible staff effort.

Finally, the system is designed to serve the needs of utilization review professionals and payers. Addition of clinician ratings to the TEaM self-report information allows the utilization reviewer to evaluate the patient's response to treatment from both the patient and therapist perspectives. The outcomes data (comparison of admission assessments with patient status later in treatment or at termination) respond to the needs of accreditation organizations and payers for evidence of treatment effectiveness.

The outcomes management model is intended to help providers perform all functions required for continuous quality improvement: initial patient evaluation, treatment planning, monitoring of patient progress, and satisfaction. To be sufficiently comprehensive while maintaining the desired brevity, the measure was designed to address nine treatment objectives: (a) Assess the severity and nature of the patient's symptoms; (b) assess the impact of the patient's problems on his or her life functioning; (c) assess patient satisfaction with treatment; (d) assess the patient's perception of the therapist ("therapeutic bond"); (e) clarify the focus (objectives) of treatment; (f) detect comorbid conditions, including chemical dependency; (g) alert the therapist to dangerous conditions (e.g., suicidality, psychosis, violence); (h) provide indicators of the reliability of information on clinical reports; and (i) monitor patient progress. Consistent with research on the assessment of the therapeutic relationship and its importance in predicting outcome (Krasner, Howard, & Brown, 1998; Saunders, Howard, & Orlinsky, 1989), a measure of the therapist-patient "bond" was included along with the clinical measures.

Measurement at the individual patient level is only one aspect of the challenge of effective outcomes management. The TEaM is designed to achieve six program

level objectives. First, use of the measurement to facilitate continuous quality improvement is a primary goal. By identifying the types of cases in which suboptimal outcomes are achieved, programs can better focus their quality improvement initiatives. Second, it is becoming increasingly important for programs to document their value in terms of patient outcomes. Thus, a second objective is to allow for the documentation of program effectiveness. Third, as the field moves toward the consistent use of evidence-based practices, the TEaM provides a method for identifying effective practices. Fourth, by combining clinical status measurement with service use data, program administrators and care managers are better able to accomplish clinically informed utilization management. Fifth, the data collected can be used to address accreditation requirements regarding the measurements of outcomes. The final program level objective is to allow for appropriate multivariate severity adjustment so that similar programs can be compared fairly.

Decisions regarding the design and development of the TEaM were based on the considerations just discussed. The remainder of this article describes the development of the TEaM instrument. Specifically, we initially describe the process by which domains and items were selected. We then report the results of a series of pilot studies intended to address issues of reliability and validity.

Development of TEaM Measures

TEaM assesses (a) a patient's subjective well-being, (b) the severity of patient symptoms associated with the most common disorders treated in outpatient settings, (c) the impact of the patient's psychological problems on his or her life (functional disability), and (d) therapeutic bond/satisfaction with treatment. These domains are grounded in extensive research on mental health outcomes. They are consistent with the Vanderbilt consensus conference (Frank et al., 1992) on outcomes of psychotherapy and are readily accepted by clinicians, operating from all major therapeutic models, as being central to clinical decisions and outcomes assessment.

The original pool of 102 scale items (excluding demographics) was constructed by Ken and the project team based on their clinical and research experience, theoretical considerations (e.g., transtheoretical model of change), and a careful reading and rewording of symptom descriptors in the *Diagnostic and Statistical Manual of Mental Disorders* (fourth edition [DSM-IV]; American Psychiatric Association, 1994).

Item selection for the final TEaM system was based on rational, theoretical selection criteria as well as empirical data analysis. Three studies were conducted to construct the scales and to ascertain their psychometric properties. On the basis of the phase model of psychotherapy, we developed measures in three main domains: subjective well-being, symptomatic status, and functional disability. These three measures were then combined to form a general measure of Behavioral Health Status (BHS).

Subjective Well-Being

The TEaM Subjective Well-Being Scale includes items about the patient's emotional and psychological adjustment and optimism about the future. We also included the 12-item Short-Form Health Survey (SF-12; Stewart, Hays, & Ware, 1988) with its two general dimensions: emotional health and physical health. This gives the scale

a broad comparability to existing studies and literature relating to outcomes research in medical settings, where the SF-12 is widely used. Thus, the initial TEaM items assessed patients' well-being in three broad areas: subjective well-being, emotional health, and physical health.

Symptoms

Based on previous research as well as current clinical nosology (i.e., *DSM-IV*; American Psychiatric Association, 1994), we devised a Symptom Checklist with subscales for seven common *DSM-IV* diagnoses: depression, anxiety, obsessive-compulsive disorder, phobia, somatization, posttraumatic stress disorder, and panic disorder. Items were also constructed to screen for substance abuse and to assess specific risk indicators and severe symptomatology that might require immediate attention (e.g., suicidal or homicidal ideation).

The stem for each item is "In the past week, how often have you . . . ? When necessary to reduce the length or reading level of an item, the *DSM-IV* statement was paraphrased (e.g., the depression symptom "diminished interest or pleasure in activities" became "felt less pleasure from things you used to enjoy." Response options are "never or rarely," "some of the time," "often," and "all or almost all of the time."

Also included in the Symptom Checklist but not contributing to its score are four items for a "fake bad" malingering index (MI) to assess the truthfulness or dependability of the patient's responses.

Functional Disability

In the Functional Disability Scale, the patient is asked to report how well he or she is doing in three major life areas: social and vocational functioning and performance of activities of daily living. The goal of the scale is to assess the extent of disability caused by the patient's psychological conditions.

The items were selected after a review of the Social Security Disability Guidelines as well as other measures of functioning. For each item, the respondent is asked to indicate "in the past week how well have you been doing in each of the following areas?" Items from the Social, Vocational, and Personal subscales, respectively, include "getting along with friends," "working accurately (making few errors)," and "making everyday decisions." Response options are "very poorly," "fairly poorly," "fairly well," and "very well."

Therapeutic Bond and Satisfaction With Care

The TEaM Therapeutic Bond Scale has three theoretical components: working alliance, understanding, and trust. Working alliance has to do with the effort the patient and the therapist put into implementing their respective roles. Understanding relates to the patient's perception that the therapist understands him or her, and trust pertains to an open, caring regard between the patient and the therapist. Using a well-established instrument in the field (Saunders, Howard, & Orlinsky, 1989) as a guide, we developed a short version that includes items from all three domains.

Patient satisfaction with the delivered care is an important component of patients' assessments of their treatment. Reviewing the literature on patient satisfaction and exploring the most widely used short version to measure the satisfaction of a

patient with the delivered care (Attkison & Zwick, 1982), we constructed items about the suitability and the progress of the treatment.

Data Quality Indexes

One of the threats to the effective use of outcome measures in clinical practice is the trustworthiness of the information contained in the measures (Lyons, Howard, O'Mahoney, & Lish, 1997). A variety of strategies exist for ensuring the ongoing reliability and accuracy of information; however, when a measure relies on self-report data, it is necessary to ensure that recipients are accurately and appropriately describing their clinical status. To facilitate the assessment of the confidence with which the findings can be interpreted, we constructed two data quality indexes: a measure of the inconsistency of responses (IRI) and a measure of malingering (MI).

The IRI and MI were developed to alert providers to the possibility that a report may not accurately reflect the patient's clinical status. This could occur when a patient is unmotivated to complete the form carefully (e.g., responds randomly), is unable to understand the form because of literacy limitations (which may be concealed because of embarrassment), or malingers his or her symptoms (e.g., in the belief that treatment costs will not be reimbursed unless he or she presents severe symptoms).

The IRI was constructed by identifying eight pairs of items and responses useful in detecting careless or random responding. For example, it is very improbable that patients would authentically report that they feel "confident about the future" almost all of the time while also reporting that they feel "hopeless or pessimistic about the future" almost all of the time. The IRI score is calculated by counting the number of inconsistent responses a patient has made; IRI scores range from 0 to 8.

The MI development was based on the "rare/preposterous symptom" method described by Rogers (1997). The MI consists of four items that might be endorsed by patients seeking to "fake bad" but are rarely reported by patients in outpatient treatment: dreams of shrinking, things crawling all over the body, others controlling movements, and awareness of strange odors that others do not notice.

Once these measurement domains were selected and item pools formed,² we conducted a set of three studies to construct and validate the TEaM Subjective Well-Being Scale, Symptom Checklist, and Functional Disability Scale and establish the reliability and validity of a composite BHS score.

Study 1: Scale Construction

The goal of Study 1 was to evaluate the item pool in each of the TEaM scales and subscales.

Method

Subjects were 600 adults in outpatient treatment. To minimize patient burden the pool of candidate items was distributed across three pilot forms. The pilot forms

²Ken had thoroughly documented his work on TEaM during the months before his final hospitalization. Unfortunately, some details of TEaM construction were inadvertently deleted from his computer on his death. For example, Ken's criteria for inclusion (exclusion) of items for the original item pool have been lost.

were designed to assess the following domains: subjective well-being, symptoms, functional disability, and therapeutic bond/satisfaction with care.

Each of the three forms was administered to 100 psychiatric outpatients at clinics in southern California and to 100 outpatients at a clinic in Rochester, New York. The forms were completed voluntarily and anonymously by patients in the waiting rooms. Patients could have been at any stage in their outpatient therapy.

Table 1 presents demographic characteristics of the TEaM development sample. The research subjects for this and the other two studies were not randomly selected, but the samples were demographically diverse and representative of patients at their respective sites. The diversity of the California and New York State sites provided for heterogeneity in the development samples. The California participants were significantly older (mean age = 44.7 years vs. 40.2; $t = 2.14$, $p < .05$), completed more years of formal education, $\chi^2(5, N = 1,526) = 42.3$, $p < .001$ (e.g., 89.5% completed "some college" vs. 36.2% of the New York State sample), and differed in its racial composition, $\chi^2(4, N = 1,526) = 20.5$, $p < .001$. Both samples were about two-thirds white, but the California sample included many Hispanics (22.8%) and few African Americans (5.3%), whereas the opposite was the case for the New York State sample (5.8% and 27.5%, respectively). There was no significant difference in gender representation.

Analyses of the pilot data focused on descriptive statistics (e.g., response frequencies) for each pilot item, and psychometric characteristics of the proposed scales and subscales of TEaM. The data were used to make an initial selection of items to be included in the next developmental stage. Therefore, we evaluated the internal consistency (alpha) for the entire scale with and without an item as well as the item-total correlations.

Results

The results of Study 1 were reviewed with two focus groups of clinicians. Item retention decisions were made on the basis of both psychometric considerations and

TABLE 1. TEaM Development Sample: Demographics for Combined Studies

Demographic variable	%	Demographic variable	%
Gender		Age	
Female	59.7	18–25	7.3
Male	40.3	26–35	21.8
		36–45	33.3
		46–55	23.6
		56+	13.9
Education		Race	
Grammar school	4.8	African American	15.2
Some high school	12.6	Asian	2.8
High school graduate	26.2	Hispanic/Latino	11.7
Some college	31.8	White	68.8
College graduate	12.0	Other	1.6
Postcollege	12.6		

Note. $N = 1,526$.

clinical utility. The resulting scales and subscales were the subject of two subsequent studies to evaluate the psychometric properties of the new scales.

Study 2: Establishing Reliability and Validity

The goal of this study was to evaluate the internal consistency reliability and the concurrent validity of the newly developed TEaM scales and subscales.

Method

Subjects were 800 adults in outpatient treatment. Concurrent validity was assessed by comparing the TEaM scores with findings from commonly used instruments that are purported to measure the same or similar constructs.

To reduce the burden on the participating patients, we again divided the questionnaire into multiple (four) packets and combined each TEaM scale (Subjective Well-Being, Symptom Checklist, and Functional Disability) with an established criterion measure assessing the same domain. As in Study 1, each of the four packets was administered to 100 psychiatric outpatients at the California sites and 100 at the New York State sites ($N = 800$ patients; $n = 200$ for each validation analysis); and the forms were completed voluntarily and anonymously by patients in the waiting rooms.

Concurrent Validity Measures

The concurrent validity of TEaM was examined by calculating the Pearson product-moment correlation between various TEaM scales and established measures of the same or similar constructs.

The 18-item General Well-Being Scale (GWS; Dupuy, 1977) is a self-report questionnaire that measures a broad range of satisfaction with self and quality of life. It is a composite measure of depressive affect, anxiety, stress, physical well-being, and sense of emotional control. The GWS has normative data on a national sample of nearly 7,000 adults and has demonstrated good reliability and validity.

The Positive and Negative Affect Scale (PANAS; Watson & Tellegen, 1985) consists of a 10-item scale of positive affect and a 10-item scale of negative affect. PANAS is based on the observation that in a number of studies of self-reported mood a positive and a negative affect factor consistently emerged as the first two varimax rotated dimensions in orthogonal factor analyses.

The Symptom Checklist-90-R (SCL-90-R; Derogatis, 1977) is the most widely used self-report measure of psychiatric symptoms. The response categories of the SCL-90-R inquire into how much distress each symptom has caused. The SCL-90-R yields nine factorially derived scores as well as three summary scores. The symptom scales are Somatization, Obsessive-Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, and Psychoticism. The SCL-90-R has norms from more than 900 nonpatients and consists of 90 items with 6 to 13 items per scale. The scales have good internal consistency and test-retest reliability.

The Beck Depression Inventory (BDI) is a self-report measure designed to assess the intensity of depression in psychiatric patients and detect depression in normal populations. The 21 items are rated on 4-point anchored response alternatives. The BDI has good psychometric properties. Beck, Steer, and Garbin (1988) summa-

rized 25 years of research with the BDI and reported a range of the internal consistency across various subject populations between .73 and .92. The test–retest coefficients ranged from .48 to .86 depending on the time intervals for retesting and sample characteristics.

The Center for Epidemiology Studies Depression Scale (CES-D; Markush & Favero, 1973) is a self-report measure with each of the 20 items being rated on a 4-point scale for severity of depressive symptoms over the previous week. It is used primarily as a screener for depressive symptoms in psychiatric, general medical, and community samples. The scale has good internal consistency, test–retest reliability, and validity.

The Social Adjustment Scale (SAS; Weissman & Bothwell, 1976) is a measure of social functioning. It surveys seven instrumental and expressive role performance areas: work, social and leisure activities, relationships with extended family, marital role and parental role, family unit role, and economic role. The items are rated on a 5-point scale; higher scores indicate more impairment. The reported psychometric properties have been adequate for this measure.

Results

The internal consistency (coefficient alpha) for the full BHS scale ($\alpha = .88$), the BHS subscales (Subjective Well-Being, $\alpha = .82$; Symptom Checklist, $\alpha = .90$; Functional Disability, $\alpha = .84$), the Bond Scale ($\alpha = .83$), and Satisfaction Scale ($\alpha = .83$) were all sufficiently high to treat each scale as a single construct.

Table 2 contains the concurrent validity comparison of the TEaM Subjective Well-Being Scale, Symptom Checklist, and Functional Disability Scale with their respective comparison measures. The correlation of both the Subjective Well-Being and Emotional Health Scales with the GWS were .76. Their correlations with the PANAS (.72 and .71, respectively) were only slightly lower. The corresponding correlations for the PH scale were much lower: .41 and .35. These findings suggested that the TEaM Subjective Well-Being and Emotional Health scales are valid indicators of psychological well-being, whereas the Physical Health Scale appeared to be assessing a separate domain of physical well-being.

TABLE 2. Pearson Product–Moment Correlations of TEaM Scales With Commonly Used Measures of Similar Constructs

TEaM scale	Comparison measure	Correlation
Subjective Well-Being	General Well-Being Scale	.76
Subjective Well-Being	PANAS	.72
Emotional Health	General Well-Being Scale	.76
Emotional Health	PANAS	.71
Physical Health	General Well-Being Scale	.41
Physical Health	PANAS	.35
Symptom Checklist	SCL-90-R GSI	.88
Depression	Beck Depression Inventory	.76
Depression	CES-Depression Scale	.89
Anxiety	SCL-90-R Anxiety Scale	.70
Functional Disability	Social Adjustment Scale	.56

Note. $N = 200$. PANAS = Postive and Negative Affect Scale; SCL-90-R = Symptom Checklist-90-Revised; GSI = Global Severity Index; CES = Center for Epidemiology Studies.

The correlation between the TEaM Symptom Checklist and the SCL-90-R total score (Global Severity Index) was .88, suggesting that the Symptom Checklist is a valid indicator of the extent and severity of psychiatric symptoms. The correlations between the TEaM Depression subscale and the CES-D and BDI scales were .89 and .76, respectively. In both cases the TEaM Depression subscale exhibited good validity. The correlation between the TEaM Functional Disability Scale and the SAS was .56. This level of validity is in the acceptable range given the diversity of the scales.

Study 3: Final Measure Construction

The primary goal of Study 3 was the assembly of the final TEaM instrument and the evaluation of its reliability and validity. A second goal was the development of a short version of the (total score) BHS scale. A third goal was to assess the validity of IRI and MI, the TEaM data quality indexes.

Method

Subjects were 126 adults in outpatient treatment. Based on the results of Studies 1 and 2, a final version of the TEaM was constructed consisting of Subjective Well-Being Scale, Symptom Checklist, Functional Disability Scale, Satisfaction With Care/Therapeutic Relationship, and MI; the IRI score is derived from items included in the other scales. To these were added demographics, items to screen for substance abuse, and severity adjustment items (e.g., chronicity, previous treatment, optimism about the future). A composite BHS score was constructed as the sum of the standardized Subjective Well-Being, Symptom Checklist, and Functional Disability scores.

The length of the instrument was reduced by deleting the Emotional Health and Physical Health Scales, because these proved redundant (Emotional Health) or of limited clinical value (Physical Health).

As in the previous studies, we collected data from a sample of cases in New York State and California ($N = 126$). We compared the TEaM with the Outcome Questionnaire (OQ-45), a 45-item self-report measure of behavioral health that includes scales for Symptom Distress, Interpersonal Relations, and Social Role (Lambert et al., 1996). The internal consistency of the Symptom Distress and Total Score are excellent (above .90). The internal consistency of the Interpersonal Relations and Social Role Scales are acceptable (.70–.75).

Results

The reliability (alpha) of the BHS scale was .88. Scores ranged from 1.6 to 5.0 ($M = 3.21$, median = 3.20, and $SD = .73$). The BHS reliability is adequate to serve its function as the primary indicator of a patient's global mental health status, which is monitored during the course of treatment.

Table 3 contains the descriptive statistics for the BHS and its subscales based on the combined samples. Comparison of the means and medians suggested that the BHS and all scales were fairly normally distributed. Among the subscales, Somatization and Panic appear to be uncommonly endorsed problems; therefore, these scales were somewhat negatively skewed.

Table 4 presents results relating to the validity of the BHS. The correlation between the TEaM BHS score and the total score on the OQ-45 was .87. The correla-

TABLE 3. Descriptive Statistics for the TEaM Scales and Subscales

Scale	Items	<i>M</i>	<i>SD</i>	Median	Range
Subjective Well-Being	3	2.89	0.93	2.80	1.0–5.0
Symptom Checklist	28	3.58	0.76	3.64	1.9–5.0
Depression	8	3.33	0.94	3.33	1.0–5.0
Anxiety	4	2.85	1.01	2.67	1.0–5.0
Phobia	3	3.54	1.10	3.67	1.0–5.0
Obsessive–Compulsive	3	3.20	1.15	3.22	1.0–5.0
Somatization	4	3.68	1.03	4.00	1.0–5.0
PTSD	3	3.69	1.11	3.67	1.0–5.0
Panic Disorder	3	3.85	1.03	4.11	1.0–5.0
Functional Disability	12	3.24	0.86	3.22	1.4–5.0
Social Functioning	4	3.25	1.06	3.34	1.0–5.0
Vocational Functioning	4	3.20	1.00	3.22	1.0–5.0
Personal Functioning	4	3.26	0.92	3.34	1.0–5.0
Behavioral Health Status	43	3.21	0.73	3.20	1.6–5.0

Note. PTSD = posttraumatic stress disorder.

tions were strongest at the most and least severe thirds of the OQ-45 distribution (.694 and .688, respectively) and relatively weak (.258) for the 33% of cases in the middle of the distribution.

On the basis of the correlations between the individual BHS items and the full-scale BHS score, we developed BHS(SF), a 15-item version of the TEaM that includes at least one item from the Subjective Well-Being Scale and from each of the Symptom Checklist and Functional Disability subscales. Generally, the latter scales were represented by the item that had the highest correlation with the BHS score. For the Study 3 sample, the correlation between BHS(SF) and the full BHS was .95; the correlation of BHS(SF) and the OQ-45 was .83.

The IRI was validated by comparing scores for the 126 Study 3 participants with IRI scores based on 50 sets of “patient responses” provided by a random number generator. In the patient sample, 78% had an IRI score of “0,” 17% scored “1,” and 5% scored “2” or higher (the flag for an “inconsistent responder”). In the random response sample, 12% scored “0,” 29% scored “1,” and 59% scored “2” or more. Thus,

TABLE 4. Pearson Product–Moment Correlations between TEaM and OQ-45 Scales

TEaM	Comparison measure	<i>r</i>
Symptom Checklist	OQ-45 Symptom Distress	.85
Functional Disability	OQ-45 Interpersonal Relations	.79
Functional Disability	OQ-45 Symptom Distress	.71
Subjective Well-Being	OQ-45 total score	.70
Symptom Checklist	OQ-45 total score	.82
Functional Disability	OQ-45 total score	.84
Behavioral Health Status	OQ-45 total score	.87
Behavioral Health Status Short Form	OQ-45 total score	.83

Note. *N* = 126. OQ-45 = Outcome Questionnaire.

the IRI identifies 5% of patients and 59% of random responders as “inconsistent.” The difference is statistically significant, $\chi^2(1, N = 176) = 67.0, p < .001$.

The MI was validated by comparing MI scores for the 126 Study 3 patients with those of 50 adults (community sample) instructed to complete the Symptom Checklist “. . . as if you were a patient seeking treatment for psychological problems. Please complete the form in a way that suggests you are mentally ill—but don’t be too obvious. (For example, don’t respond ‘all the time’ to every symptom).” Of the “fake bad” sample, 69.2% endorsed one or more of the four target items versus 50% of the patient sample, $\chi^2(1, N = 176) = 7.70, p < .01$. The most useful cutoff score for identification of possible malingerers classified 4.8% of actual patients as malingerers versus 13.5% of the “fake bad” sample, $\chi^2(1, N = 176) = 8.2, p < .01$. The sophistication of the “fake bad” sample, combined with the explicit instruction “. . . don’t be too obvious” may account for the relatively low percent of the sample that exceeded the cutoff score.

Summary and Conclusions

The findings suggest that the TEaM instrument is reliable and valid for assessing the clinical status of adults in outpatient mental health treatment. By building on the conceptual foundation developed by Ken Howard and his colleagues over the course of his career, the TEaM blends a sophisticated conceptual framework with solid psychometric properties. Given the findings of these preliminary studies, the TEaM appears to offer the potential for further advancement of the monitoring and managing of outcomes in outpatient behavioral health care. Beyond its potential for clinical research and patient assessment, the results suggest that TEaM could serve as the central component of an outcomes management system. Consistent with the design objectives, it appears that the system has the potential to be useful to researchers and clinical administrators and as a decision support tool for practicing clinicians.

The construction of an outcomes management system with TEaM as its core has begun. Four major tasks remain. First, TEaM data for a community (nontreatment) sample will be collected to establish norms for a nonpatient population, which will be useful to clinicians and utilization review professionals in determining medical necessity for treatment and in demonstrating sensitivity to pathology. Second, the administration of TEaM will be automated to provide the immediate processing and reporting of patient data essential to outcomes management. Third, longitudinal TEaM data will be collected for adults in outpatient mental health treatment. This will allow us to validate the three phase scales in relation to the sequence and rate of improvement suggested by phase theory and provide for the derivation of ETR curves, expected trajectories of recovery for individuals with different clinical characteristics seeking different forms of treatment. The longitudinal data will be used to determine the sensitivity of TEaM to phasic change. Finally, implementation research on the utility of the tool for each of its proposed applications is needed.

Although the body of Ken Howard’s research and impact on the field of psychology extends far beyond outcomes management, it is in this area that he was a true pioneer and innovator. This was also his focus at the end of his life. The TEaM instrument is among his final contributions to the field. It builds on his lifelong research contributions and provides a legacy to researchers and clinicians alike in their common goal of improving the understanding and effectiveness of mental health treatment.

References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Attkisson, C. C., & Zwick, R. (1982). The Client Satisfaction Questionnaire: Psychometric properties and correlations with service utilization and psychotherapy outcome. *Evaluation and Program Planning, 5*, 233–237.
- Barkham, M., Rees, A., Stiles, W. B., Shapiro, A., Hardy, G. E., & Reynolds, S. (1996). Dose-effect relations in time-unlimited psychotherapy. *Journal of Consulting and Clinical Psychology, 64*, 927–935.
- Barkham, M., Stiles, W. B., & Shapiro, D. A. (1993). The shape of change in psychotherapy: Longitudinal assessment of personal problems. *Journal of Consulting and Clinical Psychology, 61*, 667–677.
- Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years later. *Clinical Psychology Review, 8*, 77–100.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Derogatis, R. L. (1977). *SCL-90: Administration and procedures manual-I for the r(evised) version*. Baltimore, MD: Clinical Psychometrics Research.
- Dupuy, H. J. (1977). *A current validation study of the NCHS General Well-Being Schedule* (DHEW Publication No. HRA 78-1347). Hyattsville, MD: National Center for Health Statistics, U.S. Department of Health, Education, and Welfare.
- Frank, J. D., Luborsky, L., Wallerstein, R. S., Howard, K. I., Orlinsky, D. E., Bergin, A. E., et al. (1992). Historical developments in research centers. In D. K. Freedheim & H. J. Freudenberger (Eds.), *History of psychotherapy: A century of change* (pp. 391–449). Washington, DC: American Psychological Association.
- Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureño, G., & Villaseñor, V. S. (1988). Inventory of Interpersonal Problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56*, 885–892.
- Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-response relationship in psychotherapy. *American Psychologist, 41*, 159–164.
- Howard, K. I., Lueger, R. J., Maling, M. S., & Martinovich, Z. (1993). A phase model of psychotherapy: Causal mediation of outcome. *Journal of Consulting and Clinical Psychology, 61*, 678–685.
- Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). The evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist, 51*, 1059–1064.
- Kopta, S. M., Howard, K. I., Lowry, J. L., & Beutler, L. E. (1994). Patterns of symptomatic recovery in time-unlimited psychotherapy. *Journal of Consulting and Clinical Psychology, 62*, 1009–1016.
- Krasner, R. F., Howard, K. I., & Brown, A. S. (1998). The acquisition of psychotherapeutic skill: An empirical study. *Journal of Clinical Psychology, 54*, 895–903.
- Krause, M. S., Howard, K. I., & Lutz, W. (1998). Exploring individual change. *Journal of Consulting and Clinical Psychology, 66*, 838–845.
- Lambert, M. J. (2001). Psychotherapy outcome and quality improvement: Introduction to the special section on patient-focused research. *Journal of Consulting and Clinical Psychology, 69*, 147–149.
- Lambert, M. J., Hansen, N. B., Umphress, V., Lunnen, K., Okiishi, J., Burlingame, G. M., & Reisinger, C. W. (1996). *Administration and scoring manual for the OQ-45.2 (Outcome Questionnaire)*. Stevenson, MD: American Professional Credentialing Services LLC.
- Lueger, R. J., Howard, K. I., Martinovich, Z., Lutz, W., Anderson, E. E., & Grissom, G. R. (2001). Assessing treatment progress of individual patients using expected treatment response models. *Journal of Consulting and Clinical Psychology, 69*, 150–158.
- Lutz, W., Martinovich, Z., & Howard, K. I. (1999). Patient profiling: An application of random coefficient regression models to depicting the response of a patient to outpatient psychotherapy. *Journal of Consulting and Clinical Psychology, 67*, 571–577.
- Lyons, J. S., Howard, K. I., O'Mahoney, M., & Lish, J. (1997). *The measurement and management of clinical outcomes in mental health*. New York: Wiley.
- Maling, M. S., Gurtman, M. B., & Howard, K. I. (1995). The response of interpersonal problems to varying doses of psychotherapy. *Psychotherapy Research, 5*, 63–75.
- Markush, R. E., & Favero, R. V. (1973). Epidemiologic assessment of stressful life events, depressed mood, and psychophysiological symptoms: A preliminary report. In B. S. Dohrenwend (Ed.), *Stressful life events: Their nature and effects* (pp. 171–190). New York: Wiley.
- Martinovich, Z. (1998). *Evaluating a phase model of psychotherapy outcome: An application of hierarchical logistic modeling*. Unpublished doctoral dissertation, Northwestern University, Chicago.
- Pilkonis, P. A., & Frank, E. (1988). Personality pathology in recurrent depression: Nature, preva-

- lence, and relationship to treatment response. *American Journal of Psychiatry*, 145, 435–441.
- Rogers, R. (1997). *Clinical assessment of malingering and deception*. New York: Guilford Press.
- Sangslund, S. (2001). Chemical dependency outcomes management. In K. Coughlin (Ed.), *Behavioral outcomes & guidelines sourcebook* (pp. 312–320). New York: Faulkner & Gray.
- Saunders, S. M., Howard, K. I., & Orlinsky, D. E. (1989). The therapeutic bond scales: Psychometric characteristics and relationship to treatment effectiveness. *Psychological Assessment*, 1, 323–330.
- Sperry, L., Brill, P. L., Howard, K. I., & Grissom, G. R. (1996). *Treatment outcomes in psychotherapy and psychiatric interventions*. New York: Brunner/Mazel.
- Stewart, A. L., Hays, R. D., & Ware, J. E. (1988). The MOS short-form General Health Survey. *Medical Care*, 26, 724–735.
- Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98, 219–235.
- Weissman, M. M., & Bothwell, S. (1976). Assessment of social adjustment by patient self-report. *Archives of General Psychiatry*, 33, 1111–1115.

Zusammenfassung

Seine Studien zu Ablauf (Phasenmodell), Raten (Dosierungs-Reaktionsmodell) und Vorhersagbarkeit (Modell zur erwarteten Therapiereaktion) von Veränderungen, die im Verlauf von Psychotherapie auftreten, gehören zu Ken Howards wichtigsten Beiträgen zur Psychotherapieforschung. Diese Forschung bildete die Grundlage für die Entwicklung eines Erfolgsmanagementsystems: Behandlungsevaluation und -management (Treatment Evaluation and Management [TEaM]). Die Autoren beschreiben die Entwicklung von TEaM-Massen entsprechend der zentralen Konstrukte des Phasenmodells: subjektives Wohlbefinden, Symptome und funktionale Beeinträchtigungen. Diese Scores wurden zu einem allgemeinen Maß zusammengefügt, dem Index zum Behavioral Health Status (BHS), der ein globales Maß des psychologischen und Funktions-Status eines Patienten darstellt. Die Autoren geben Informationen zu den psychometrischen Indices der einzelnen Masse und dem aus ihnen zusammengesetzten BHS-Index. Ken Howard hat kurz vor seinem Tod noch an der Optimalisierung von TEaM gearbeitet. So kann dieser Messansatz als sein letzter Beitrag im Bereich des Erfolgsmanagements gesehen werden.

Résumé

Les contributions majeures de Ken Howard à la recherche en psychothérapie comprennent ses études de la séquence (modèle de phases), du taux (modèle dose-réponse), et de la prédictibilité (modélage de la réponse au traitement attendue) de changements arrivant au cours de psychothérapies. Ce corpus de travail a formé le fondement pour le développement d'un système de management des résultats : évaluation et management du traitement (TEaM). Les auteurs décrivent le développement de mesures TEaM correspondant aux construits centraux du modèle de phases : bien-être subjectif, symptômes et trouble fonctionnel. Ces scores sont combinés pour en dériver un score composé, l'index du Behaviour Health Status (BHS) qui livre une mesure globale de l'état et du fonctionnement psychiques d'un patient. Les auteurs présentent de l'information psychométrique au sujet de chacune des mesures et de l'index composé BHS. Ken était en train de travailler à la terminaison de TEaM à l'époque de sa mort. Ainsi, cette approche de mesurage est sa contribution ultime au champ du management des résultats.

Resumen

Las principales contribuciones de Ken Howard a la investigación en psicoterapia incluyen sus estudios de la secuencia (modelo de fase), relación (rate) (modelo dosis-respuesta) y predictibilidad (modelo de la respuesta esperada al tratamiento) de los cambios que ocurren durante una psicoterapia. Este conjunto de trabajos constituyó las bases para el desarrollo del sistema de administración de resultados: evaluación y administración del tratamiento (TEaM). Los autores describen el desarrollo de medidas TEaM correspondientes a los constructos centrales del modelo de fase: bienestar subjetivo, síntomas y discapacidad funcional. Estos puntajes se combinan para obtener un puntaje compuesto, el índice del Status de salud comportamental (BHS), que da una medida global del status y el funcionamiento psicológico del paciente. Los autores presentan una información psicométrica de cada medida y del índice compuesto BHS. En el momento de su muerte, Ken trabajaba en la conclusión del TEaM. Así, este enfoque de medición fue su última contribución al campo de la administración de resultados.

Resumo

A grande contribuição de Ken Howard para a investigação em psicoterapia inclui os seus estudos da sequência (modelo das fases), ritmo (modelo dose-resposta) e previsibilidade (modelagem da resposta esperada ao tratamento) das mudanças que ocorrem em psicoterapia. Estes trabalhos formam os alicerces do desenvolvimento de um sistema de gestão dos resultados: Gestão e Avaliação do Tratamento (TEaM). Os autores descrevem o desenvolvimento das medidas do TEaM correspondentes aos constructos centrais do modelo das fases: bem estar subjectivo, sintomas, e perturbação funcional. Estes resultados são combinados de modo a derivar um resultado composto, o índice de Estado de Saúde Comportamental (Behavioral Health Status, BHS), que fornece uma medida geral do estado psicológico e funcionamento do paciente. Os autores apresentam informação psicométrica de cada medida e do índice composto BHS. Ken Howard estava a trabalhar na finalização do TEaM na altura da sua morte. Portanto, esta medida é a sua última contribuição para a área da gestão do resultados em psicoterapia.

Sommario

I più grandi contributi offerti da Ken Howard alla ricerca psicoterapica includono i suoi studi eseguiti sulla sequenza (modello fase); percentuale (modello somministrazione-risposta), e prevedibilità (modello risposta attesa al trattamento) dei cambiamenti che hanno luogo durante la psicoterapia. Questa mole di lavoro ha costituito le fondamenta che hanno portato allo sviluppo di un sistema di gestione degli esiti: valutazione e gestione del trattamento (TEaM). Gli autori descrivono lo sviluppo delle misure TEaM che corrispondono alle strutture centrali del modello di fase: benessere soggettivo, sintomi, disabilità funzionale. Questi punteggi sono combinati in modo tale da ottenere un punteggio composto, l'indice di Stato Comportamentale (BHS), il quale fornisce una valutazione generale della condizione e della funzionalità psicologica del paziente. Gli autori presentano informazioni psicometriche relative ad ognuna delle valutazioni e all'indice composto BHS. Ken stava lavorando al perfezionamento del TEaM, al momento della sua morte. Risulta quindi, che questo approccio di misura è il suo ultimo contributo al settore della gestione degli esiti.

Received July 20, 2001

Revision received July 1, 2002

Accepted July 1, 2002