Clinicians and Outcome Measurement: What's the Use?

Ann F. Garland, PhD
Marc Kruse, BA
Gregory A. Aarons, PhD

Abstract

The goal of this study was to learn more about clinicians' experiences with, and perceptions of the utility, validity, and feasibility of standardized outcome measures in practice. Fifty randomly selected clinicians from multiple disciplines and multiple service agencies in a large children's public mental health service system were interviewed individually (n = *30) or in focus groups* (n = *20) using semistructured interviews. There was great variability across clinicians in attitudes about empirical methods of treatment evaluation. There was consensus regarding feasibility challenges of administering standardized measures, including time burden and literacy barriers. Although all participants had received scored assessment profiles for their clients, the vast majority reported that they did not use the scores in treatment planning or monitoring. Their suggestions for improved clinical utility of outcome measurement are included. With increased attention and resources devoted to performance outcome assessment, it is concerning that most clinicians perceive little clinical utility of outcome measurement.*

Mental health practitioners are increasingly being encouraged to adopt evidence-based practices, including empirically supported treatment and assessment methods. Evidence-based practice emphasizes the systematic empirical evaluation of outcomes. Thus, one of the major developments in the delivery of mental health services over the last several years has been the increased pressure on providers to collect standardized outcome data on clients served. Such data are now required by many public and private agencies,[1] and quality of care requirements from major accreditation organizations such as the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) include collection of outcome data.[2] Despite all of the research, administrative, and policy attention to outcome measurement in mental health services, the actual clinical utility of outcome measurement remains largely unexamined.

There are few studies examining clinicians' attitudes and behaviors regarding the use of outcome measurement in practice.[3] However, there is likely great variability in the extent to which clinicians value empirically supported assessment and/or intervention practices.[4] Unless mandated, most clinicians are not likely to use standardized measures to assess clients, nor to empirically evaluate progress

---

Address correspondence to Ann F. Garland, PhD, Associate Director, Child and Adolescent Services Research Center, and Associate Professor in Psychiatry, University of California, San Diego, 3020 Childrens Way (MC 5033), San Diego, CA 92123. E-mail: agarland@casrc.org.

Marc Kruse, BA, was Research Assistant at the Child and Adolescent Services Research Center and is now in the Department of Psychology at the University of Texas, Austin.

Gregory A. Aarons, PhD, is Research Scientist at the Child and Adolescent Services Research Center and Assistant Clinical Professor in Psychiatry at the University of California, San Diego.

Journal of Behavioral Health Services & Research, 2003, 30(4), 393–405.

in treatment.[3,5] Many clinicians may view outcome assessment as cumbersome and/or intrusive.[5] Given this context, it is important to delineate how newly mandated requirements to collect standardized outcome data have affected clinicians' attitudes and practices. Understanding more about clinicians' attitudes and practices will improve our understanding of the context in which data are being collected and the potential impact on clinical care.

Outcome assessment in mental health is becoming ubiquitous in many settings, yet the positive impact on quality or effectiveness of care is more theoretically than empirically based. There is limited evidence that implementation of outcome assessment protocols, in and of itself, improves the overall effectiveness or quality of mental health services.[6,7] However, there is a great deal of presumed benefit and potential utility of outcome assessment at two levels: (1) at the aggregate level and (2) at the individual clinician level. At the aggregate level, standardized outcome data may have utility for program planning, funding decisions, and more rigorous monitoring of quality of care across systems or agencies. The utility of aggregate data has general acceptance as an evaluative strategy, and has been shown to improve medical services.[8] Support for individual clinical utility is derived from studies demonstrating the fallibility of clinicians' anecdotal judgment, compared to standardized, "actuarial" data.[9–11] These studies support the use of standardized assessment tools over clinicians' judgment for clinical assessment and decision-making. The focus of the present study is the perceived utility of outcome measurement for individual clinicians.

A naturalistic experiment in California provides the context for this study of clinicians' attitudes about, and experiences with, outcome measurement. In 1994, the State Legislature of California passed a law requiring providers of institutionally based, publicly funded mental health services to collect, and submit to the state, performance outcome measures at intake and subsequent follow-up intervals.[12] This Performance Outcome Project (POP) is a good example of many similar outcome assessment protocols for children, in that the required measures (eg, Child Behavior Checklist, CBCL[13], Youth Self-Report, YSR,[14] Child and Adolescent Functional Assessment Scale, CAFAS,[15] Client Satisfaction Questionnaire, CSQ,[16] and time-frames are consistent with other such protocols nationally.[17]

The POP mandate did not include additional funding for collection of outcome measures, but rather, the administration of measures was to be subsumed under standard clinical practice. State officials surveyed county representatives to estimate the clinical and clerical costs of collecting and submitting data (excluding equipment and materials costs). Estimates ranged from $15 to $62 per client, per year (Sala-Moore, PC, March 2000). Given that the annual target population is approximately 60,000 children, these costs for POP could range from $900,000 to $3,720,000 each year. While these are rough estimates, the point is that implementation of outcome assessment protocols requires significant personnel, equipment, and supply resources. This does not include the more subjective "costs" to staff morale and administrative burden. Given these objective and subjective costs, it is critically important to examine the utility of such outcome assessment initiatives.

Successful adoption (and adaptation) of evidence-based practices for community settings requires greater attention to the contextual factors that may facilitate or impede such adoption.[18,19] One such set of factors that has received minimal research attention is clinicians' attitudes and concerns regarding the adoption of evidence-based practices.[20,21] The limited available studies do suggest that there are significant attitudinal barriers among many clinicians.[20,21] These barriers may include performance evaluation concerns, logistical/feasibility considerations, and conceptual appropriateness of measurement tools.[20] Improved understanding of the content and range of clinicians' attitudes and perceptions, and the extent to which these may be barriers or facilitators for the dissemination and adoption of empirically based clinical methods, is necessary. Greater understanding of the cultural context of practicing clinicians in community-based sites is essential for efforts to ultimately improve practice.

The goal of this study was to learn more about clinicians' experiences and perceptions of the utility, validity, and feasibility of outcome measures, and their suggestions for improved methods. Qualitative and quantitative data collection methods were used to address the following questions: (a) To what extent do clinicians believe it is possible to quantitatively measure the effectiveness of treatment? (b) How do clinicians evaluate their own effectiveness? (c) Do clinicians feel increased pressure to quantitatively demonstrate their effectiveness? (d) To what extent and how do clinicians use the standardized outcome data mandated by the state? (e) What are the reported barriers to the use of the standardized outcome data? (f) What changes would clinicians like to see in outcome evaluation? (g) Are there differences in attitudes or experience regarding outcome assessment by discipline of clinician?

Methods

Participants

Participants were 50 clinicians from San Diego County who had participated in the California State mandated POP assessment since inception of POP data collection in July 1997; interviews for this study were conducted in 1999. All participants had attended a county-mandated 4-hour training session on the use and interpretation of the required POP measures (CBCL, YSR, CAFAS, CSQ). The 358 eligible clinicians came from mental health agencies that receive public funding (MediCal) in the county of San Diego. Clinicians were randomly selected for recruitment until complete data were obtained from 50 participants. Recruitment for study participation was attempted for 117 clinicians. Reasons for nonparticipation among the 67 nonparticipants included the following: 28 (42%) never returned repeated phone calls, 20 (30%), were contacted, but were unable to schedule an interview (eg, most cited time constraints), 12 (18%) did not feel they would be able to provide useful information (eg, not enough experience), 3 (4%) agreed to participate but later canceled, did not show, and/or reschedule, and 4 (6%) directly declined participation.

Table 1 lists descriptive data about the 50 study participants compared to data from a much larger inclusive survey of 284 clinicians from the same county mental health system.[22] The clinicians participating in this study appear to be generally representative of the clinicians across the county in terms of age, gender, education, mental health discipline, years of experience, and service settings. The majority were female (80%) and Caucasian (76%). Participating clinicians ranged in number of years experience in the mental health field from 1 to 37 ($M = 12.8$, $SD = 8.4$) years, and in age from 24 to 62 ($M = 39.7$, $SD = 9.5$). The majority of participants had master's level degrees (62%); social work and counseling were the most common disciplines represented. The participants were recruited from 20 different programs; the majority worked in outpatient clinics (62%). All participants reported that they provided psychotherapeutic services to youths and families. Regarding primary theoretical orientation, 28% of the clinician participants endorsed family systems, 26% eclectic, 24% cognitive behavioral or behavioral, 16% psychodynamic, 2% humanistic, and 4% "other."

Measures and procedure

Clinician interview

A semistructured interview was developed by the investigators to examine clinicians' attitudes and experiences regarding the evaluation of outcomes of youth mental health services in individual interview and focus group formats.* Both formats were utilized to maximize the variability in responses and minimize the effect of response biases. For example, individual interviews rely heavily on the individual's recollection of events, whereas the focus group format might prompt for memories of experiences. Alternatively, group dynamics might bias and/or intimidate some informants who might share attitudes more openly in an individual interview. Participants were assigned to individual interview or focus group on the basis of a goal of diversifying the composition of the focus groups (by discipline, agency affiliation, gender, and race/ethnicity), and by scheduling availability.

---

*Copies of the interview and the coding scheme are available upon request from the authors.

Table 1
Study participants vs "population"

	Current study ($N = 50$)	Aarons[20] ($N = 284$)
Age	39.7 (9.5)	35.7 (10.27)
Gender		
Male	20	23
Female	80	77
Ethnicity		
Caucasian	76	65
Hispanic	2	15
African American	6	7
Asian/Pacific Islander	10	6
Other	6	7
Education level		
No degree	NA	4
BA, BS	11	21
Some graduate school	NA	10
MSW, MA, MFT	66	55
PhD, MD	23	10
Years experience	12.8 (8.4)	8.2 (7.7)
Primary discipline		
Counseling	24	32
Social work	32	31
Psychology	40	23
Psychiatry	0	2
Other	4	12
Services provided		
Outpatient treatment	62	49
Day treatment	16	19
Assessment/evaluation	0	10
Case management	10	8
Residential treatment	6	6
Other	6	8

Note: All values are expressed in percentage (%) except 'Age' and 'Years Experience' for which mean and standard deviation are given.

Focus groups were conducted at the research office, whereas most of the clinician interviews were conducted in the clinicians' offices.

Questions for the interview were generated by the authors on the basis of study goals and pilot discussion with several clinicians about their experiences with outcome measurement; the interview questions were the same for both formats. The format and order of the questions were informed by consultation with experts in qualitative data collection (Drs Ed McQuarrie and Rae Newton). The semistructured interview was pilot tested with 5 clinicians prior to implementation; minor revisions were made for clarification.

The interview began with basic questions about the clinicians' work settings and patient population and progressed toward more sensitive issues regarding their perception of various methods of evaluating the effectiveness of their work. The clinicians were then asked questions assessing their perceptions of the utility, validity, and feasibility of standardized outcome measures and how they as clinicians have been affected by increasing pressure to quantitatively document the effectiveness of mental health services. Finally, each clinician was asked to suggest improvements in methods of evaluating the effectiveness of services.

Clinicians' self-report questionnaire

Following completion of the individual interviews and focus groups, each participant completed a brief questionnaire. The questionnaire contained basic descriptive items including gender, race-ethnicity, age, education level, number of years involved in the mental health field, primary discipline, primary therapeutic orientation, and primary work setting. In addition, each clinician was asked to indicate the extent to which they used various methods for evaluating the effectiveness of mental health services for adolescents in their clinical practice. They were also asked to indicate the relative importance of the 5 outcome domains identified by Hoagwood and colleagues[23] (ie, Symptoms, Functioning, Consumer Perspectives, Environment, and Systems) using a Likert-type scale, with responses ranging from 1, "Not at all Important," to 4, "Very Important." Operational definitions and examples for each outcome domain were taken from the original model by Hoagwood and colleagues.[23(Table 1, p1059)]

Qualitative data analysis

Each interview and 2 of the 3 focus groups* were audio-recorded and later transcribed for coding. The purpose of the coding was to assign unique labels to text passages that contained references to specific categories of information; text passages could be of any length, but were defined as communicating one idea. Examples of typical text passages are provided in the results section. The data were coded using QSR Nudist software to examine the frequency of specific categories of responses.[24] The systematic coding scheme was developed by the authors on the basis of the range of possible responses. Each author read a random sample of at least 5 different transcripts to generate the possible response categories. A comprehensive list of response options was then pilot-tested by each of the authors on one interview in order to determine the appropriateness of the coding scheme for the subject material and ensure that different coders could independently assign the same codes to the same material.

After independent pilot-testing, the authors met to discuss the appropriateness of the coding scheme and to examine the degree of agreement between coders. The coding scheme was refined to reduce redundant codes and vague definitions. Each transcript was then coded independently by 2 of the authors. Interrater reliability was assessed for the first 10 coded interviews. For all coded text segments, the coders agreed on the codes over 70% of the time. Coding discrepancies (eg, 1 coder attached a code to a particular passage when the other did not, or 2 coders attached 2 different codes to the same passage) were discussed until the 2 coders could arrive at an agreement.

All responses within categories corresponding to the central study questions were printed and representative, concise quotes were selected to reflect the responses in that category. These quotes are used in the results section below to exemplify the descriptive summary of the qualitative data.

Results

Results are summarized below in subsections addressing the central questions of the study.

Do clinicians believe it is possible to quantitatively measure the effectiveness of treatment?

In general, there was great variability in clinicians' attitudes about the extent to which it is possible to quantify the effectiveness of treatment. This variability is reflected in the range of responses across all clinician respondents, as well as the variability (or ambivalence) expressed by individual

---

*The first focus group was not transcribed because of technical difficulties that resulted in an unintelligible recording. Only self-report data from these participants ($n = 4$) are included in the results reported here.

clinicians. Across clinicians, the percentage of respondents who indicated that it was not possible to quantitatively measure change in treatment was roughly equal to the percentage who indicated that it was possible. Those who stated that it was not possible ranged from those who were ideologically strongly opposed to quantifying the complexity and nuance of human change in psychotherapy (approximately 25% of participants), to those who felt that the measurement of psychotherapy outcome is virtually impossible (also 25%). The following quote exemplifies the ideological opposition: *"The idea that you can take human behavior and objectify it is only one paradigm and it is an extremely limited paradigm...At the point where we try to take these things and objectify them, then you have destroyed them already because you are taking something and trying to put it into a language that it just won't fit."* Strong skepticism about efforts to measure change in therapy was expressed as follows: *"They're trying to make this like an exact science like other fields, but it's not. It's not very accurate; it's never going to be very accurate."* Alternatively, a few respondents expressed more confidence or optimism regarding measuring change in treatment: *"I guess it started with my training, but I have always believed that the best way to identify changes in kids is by looking at the child's measurable behavior."*

Although some clinicians expressed clearly positive or negative attitudes regarding the quantitative measurement of treatment effectiveness, many expressed ambivalence. Such responses reflected general support for attempts to measure outcomes, but frustration or recognition of the challenges inherent in measuring these constructs. The following quote nicely illustrates a commonly expressed sentiment: *"It's like picking up jello with a fork because there are so many variables. I don't think it's easy at all to quantify these things, but I think, yes, it should be done."*

How do clinicians evaluate their effectiveness?

Clinicians were asked to rate the frequency with which they use each of 7 potential methods of evaluating their clinical work. The response options ranged from 1 "Not at All" to 4 "Very Often" and the mean ratings for each method are listed in Table 2, below, in descending order of frequency of use.

Table 2 indicates that clinicians endorse the use of a variety of methods for evaluating the effectiveness of their work and there is little variability in the reported frequency of use of many different methods. A notable exception is the reported use of standardized measures, which is much less frequent (rating of 2.12 compared to range of 3.22–3.54 for all other methods).

Data from the qualitative interviews is generally consistent with the scaled questionnaire ratings. The most commonly mentioned methods of evaluating effectiveness of treatment included subjective reports from the client, followed closely by reports from parents and teachers. A large majority of respondents also reported using clinician observation and intuition as their evaluative method. One participant articulated this as follows: *"There is a sense you can get from their thoughts, their psyche."* Fewer than 10% of participants mentioned use of any standardized measures or scales

Table 2

Clinicians' reported frequency of use for methods of evaluating effectiveness ($n = 50$)

Method	Mean rating	SD
Real world functional indicators (eg, school grades, disciplinary actions)	3.54	0.68
Clinician intuition/feelings	3.46	0.73
Progress toward individualized treatment goals	3.44	0.76
Other informants' reports of client functioning	3.40	0.78
Client self-report of symptoms/functioning	3.34	0.77
Client reported satisfaction with treatment	3.22	0.84
Standardized measures or scales	2.12	0.82

and no respondents spontaneously mentioned the specific state mandated standardized outcome assessment measures.

Clinicians were also asked to identify the motivation and/or incentives to evaluate the effectiveness of their work and the majority mentioned mandatory participation in interdisciplinary "treatment team" case review mechanisms built into the structure of their settings. Many also mentioned structural, funding-based requirements to outline treatment goals at the start of treatment and review progress toward those goals over the course of treatment. Endorsement of this type of practice varied dramatically; several respondents voiced strong opposition, or at least cynicism, about the definition and review of treatment goals. For example: *"as far as these form things go, with these goals and all of these objectives, those are just trashed at; it's to satisfy some requirement, so we'll put something down."* Others' comments reflected more investment in the process; *"every time we see a client we're focusing on the goals of treatment and we review their goals and if they need to be changed, we'll change them."* Only a couple of respondents mentioned that any evaluation of effectiveness was self-motivated or implemented to improve their own practice.

Clinicians also rated the importance of 5 outcome domains for evaluating a client's progress in treatment. The outcome domain categories were adapted from Hoagwood and colleagues' conceptual model of outcome indicators for youth services.[23] The response options ranged from 1 "Not Important" to 4 "Very Important" and the mean ratings for each domain are listed in Table 3, in descending order of rated importance. Improved functioning received the highest "importance" rating and client satisfaction the lowest, but there was not a great deal of difference between these two (ratings of 3.80 and 3.22 respectively).

Responses from the open-ended qualitative interview were generally consistent with the ratings displayed in Table 3. Virtually all respondents (90%) stated that improvement in the home environment (eg, family dynamics, parenting practices) was a desired outcome indicator. Other frequently noted outcome indicators included decreased symptoms (85%), cognitive changes (eg, improved understanding of problems, ability to identify/express emotions) (85%), improved school functioning (80%), improved home functioning (eg, compliance) (80%), improved social functioning (80%), and changes in general mood or affect (70%). Strength-based outcomes (eg, improved skills) were mentioned by 60% of the respondents and individualized client-specific changes by 50%. Less frequently stated indicators were school environment changes (45%) and improved community functioning (decreased drug use and/or gang involvement) (20%).

Do clinicians feel increased pressure to quantitatively demonstrate their effectiveness?

Most clinicians reported that they did not experience increased pressure to demonstrate their effectiveness as a psychotherapist (although the majority felt pressured to complete paperwork as noted below). Within this group, some comments reflected perceived immunity to such pressure based on confidence in clinical abilities: eg, *"I feel zero pressure . . . because I am a good clinician and I know that I am a good clinician, and my clients know that I am a good clinician."* However the majority of comments on this topic reflected the perception that outcome data are not being used

Table 3

Clinicians' reported importance for outcome domains ($n = 50$)

Outcome domain	Mean rating	SD
Improved functioning	3.80	0.45
Symptom reduction	3.76	0.59
Environmental stability	3.50	0.68
Decreased service use	3.24	0.77
Client satisfaction	3.22	0.71

to evaluate clinician effectiveness, and therefore there is no individual accountability pressure. *"I don't feel pressure, because nobody looks at the measures"*; *"People haven't felt too much pressure, because they haven't seen how it is being used."*

A sizable minority of clinicians (approximately one third) did report experiencing increased pressure to document effectiveness, but this was usually expressed as increased pressure on all mental health professionals, not specific individual pressure. For example, *"mental health has always had a negative image in the public eye anyway. We've had a lot to prove; so to prove that therapy even has a benefit to these kids, yes, I think there is a huge push to prove the value of this field to politicians and society."* Very few clinicians indicated that they felt direct personal pressure to demonstrate the effectiveness of their services and/or to demonstrate improved clinical performance.

To what extent do clinicians use the standardized outcome data mandated by the state?

The vast majority of respondents (92%) indicated that they had never used the scores from the mandated standardized measures in their clinical practice (eg, for diagnostic evaluation, treatment planning, or treatment monitoring). The reported barriers to both the collection and utilization of these measures are summarized in the next section. However, responses to this topic also reflected ambivalence. Specifically, although almost every respondent stately directly that they did not use the scores from the measures in their practice, slightly more than half (60%) of the respondents did make comments in the interviews reporting how the process of administering the measures to the youth and parent was useful. The most common example of utility involved comparing the youth and parent responses, eg, *"What does intrigue me is reading what a mother knows about her kid and what the kid knows about him or herself... because it shows a lot about their relationship."* Several clinicians reported that the process of collecting the instruments was a useful step in the intake process; eg, *"For the parents, I think it is a positive experience because they feel that they have not had an opportunity in the past to provide their input and I think it helps them think about what their child is going through"* and *"Clinically, I can find it helpful even in just giving it to them because it can sometimes open the door for me to talk about something... sometimes it does give me an answer to something that I would have otherwise found out too late."*

What are the reported barriers to use of the standardized outcome data?

The reported barriers to the use of standardized outcome data fall into the following 3 general categories: (a) feasibility concerns, (b) perceived invalidity, and (c) interpretation difficulties. Regarding feasibility, almost all respondents (90%) indicated that there was a significant time burden for collecting the data and that this time burden was greatest for the clinician, compared to the parents, youth, or other staff members. The following quote is representative of the sentiments of the vast majority of the clinicians: *"As clinicians, we are burdened by more and more paperwork to prove what we do, which takes a big hunk of time away from doing what we do."* Almost all clinicians expressed frustration regarding the time it takes to administer the forms. There was great variability in reports of actual time demand, ranging from approximately 20 to 90 minutes per informant (youth and parent), plus clinicians' own time to rate the CAFAS, which ranged from approximately 10 to 60 minutes.

Many clinicians (45%) also indicated that the instruments are very difficult for parents and/or youth to understand and a minority (20%) felt that the experience of completing the forms was very stressful for their clients and families. Sixty percent of clinicians reported that they had received negative feedback from parents and youth about the process of completing the forms, whereas only 10% reported receiving positive feedback.

The clinicians also raised concerns about the perceived validity of the standardized outcome assessment measures. Many (55%) believed that the measures were not appropriate, nor valid, for their particular patient population. The majority of these cited cross-cultural issues; eg, *"this whole thing is more geared towards the Anglo population. I don't think it's very culturally sensitive... it's not well written for the population we serve."* Other participants indicated that the measures were *"too behaviorally oriented,"* and did not focus on *"internal processes"* or other constructs most important to them, such as parent-child relations. A sizable percentage of the respondents expressed skepticism about the validity of the scores on the measures because of their perception that many of the youth and parent informants did not understand the forms and/or may have had response biases that affected the results.

Another barrier in reported use of the standardized data was the clinicians' reported difficulty interpreting the meaning of the scores. Several clinicians indicated that they were not sure if the measures validly reflected changes in their clients' functioning because they did not read (25%) or understand (15%) the results. Many felt that the feedback they received, which included a standard profile of the CBCL and YSR scores,[25] was not "user friendly." Even most of those who reported that they understood the feedback, indicated that they didn't find the scores helpful in practice. For example: *"at this point looking at the scores isn't a priority . . . so, I don't even look at them anymore and I don't think anybody else does."* Several respondents indicated that the scores offered them no new information: *"I've never seen anything on a printout that I didn't already know"; "I don't think I need the scores to tell me if the child has improved. I'm in touch with the child and the progress of the child in more significant ways."*

What changes would clinicians like to see in outcome evaluation?

The participants were asked what different types of constructs or variables should be included in outcome evaluation. Approximately one third of the respondents suggested client history data, such as presence of learning disabilities, previous service use, and maltreatment experience. Slightly fewer suggested family-level variables such as parental history of psychopathology and "cultural issues." Finally, a minority (10%) mentioned assessing environmental variables such as poverty and gang involvement.

The most common suggestions for improving the methods of outcome evaluation involved improving the feasibility of the measures and simplifying the interpretation of the scores. Specifically, approximately half of the respondents requested instruments with briefer administration and simpler language. Half also requested simpler graphic and narrative interpretations of the data. Many clinicians indicated that they would be more likely to use scores from standardized assessment measures if the results were presented in narrative, as opposed to quantitative form. Some gave the desired example of a summary in a psychological testing report.

Many respondents requested more training and information on how to use interpretation of the scores in practice. The following comments reflect the sentiments of many respondents: *"One of the things I'd like to see is for someone to convince us therapists about why we need to do this . . . What benefit is this to me and what can I get off of this that would really be helpful to me as a therapist?"* Several participants suggested that clinicians should have been involved in planning the outcome assessment protocol from the start. For example, *"Why these measures? Why weren't we involved in the first place in the decisions?"* Additional comments reflected a similar theme of lack of professional respect for the clinicians in this process; *"We're doing a lot of extra work for this and we're expected to do it with no questions asked . . . The rewards are so minimal, which does not feel professionally respectful nor supportive of me or anybody else I know who is doing it."*

Are there obvious differences in attitudes or experience regarding outcome assessment by discipline of clinician?

There was a fairly even representation of psychologists, social workers, and counselors included in the study and exploratory analyses were conducted to identify any obvious group differences. Analyses of the self-report frequency data for endorsement of different evaluation methods or preferred outcomes (Tables 1 and 2) did not reflect any significant, nor trend, group effects by discipline. The interview data similarly revealed no obvious differences in attitudes or experience, with one exception. A few of the psychologists were interns or postdoctoral fellows and these individuals tended to offer the most positive comments regarding the utility of the standardized outcome assessment measures. Likewise, they tended to be the least likely to report difficulty interpreting the scores of the measures. However, this was not generalized to staff psychologists, whose comments were not significantly different from the range of comments by social workers and counselors.

Discussion

The clinician participants in this study were very willing to discuss their attitudes about and experiences with outcome measurement and the evaluation of their work. The seriousness with which most approached the task is reflected in the thoughtfulness and depth of many remarks. Many participants expressed appreciation for the opportunity to share their opinions and experiences. The data reflect rich complexity and variability in attitudes and experience. This variability is clear across respondents, but also within respondents, suggesting that many have ambivalent opinions about some of the issues raised here. The qualitative data collection method highlighted the complexity and the ambivalence in attitudes since it allowed responses reflecting contrasting opinions. This variability in clinicians' receptivity to empirically based methods of outcome assessment is quite consistent with Addis and Krasnow's report of variability among psychologists in receptivity to empirically based treatments.[4]

Variability in clinicians' attitudes about outcome measurement was reflected most dramatically in their response to the broad, initial question about the extent to which it is possible to quantitatively measure the effectiveness of treatment. Responses ranged from clear ideological opposition of empirical assessment of effectiveness, to skepticism about our current methods of assessment, to full support for empirical assessment. The majority of respondents gave responses falling in more than one of these categories, reflecting ambivalence. While the majority of participants supported the goal of measuring outcomes, they also stressed the complexity and difficulty in meeting this challenge.

Clinicians cited many different preferred sources or methods for evaluating the effectiveness of their work in standard practice; the most common were subjective reports from the client and/or family members and teachers, and their own (clinicians') observation and intuition. On a self-report form, they rated the use of clinician intuition very highly, along with tracking of functional indicators (eg, school grades, disciplinary actions). The strong endorsement of clinician intuition as an evaluative strategy stands in contrast to the lack of endorsement of the use of standardized measures. Both the open-ended interview responses and the self-report data confirm that these clinicians rarely use any standardized measures or scales to evaluate clients' progress in treatment. The lack of use of measures in standard practice is consistent with others' findings.[3,5] One difference here is that all clinicians had participated in a mandated outcome assessment protocol, and so they all had experience with this evaluative strategy. This universal experience did not necessarily result in endorsement of the practice.

Clinicians' anecdotal observations and intuitions are subject to many perceptual biases and have been shown to be less reliable and valid than actuarial data from standardized measures.[9–11] Thus, it is concerning that clinicians report being much more likely to evaluate the effectiveness of treatment by using their own intuition, compared to any standardized measure. These findings are not surprising however, given the body of evidence suggesting that among mental health clinicians, personal beliefs are usually more influential than scientific evidence.[10] Beutler has argued cogently that the field needs better standards of effectiveness than those based on "good intentions and strong beliefs."[26(p999)] The current study suggests that most clinicians appreciate and support the need for empirically based effectiveness evaluation, but they are frustrated by currently available methods, which many perceive to be unfeasible, invalid, and lacking in clear clinical application.

The reported barriers to clinicians' use of standardized outcome measures included frustrations with feasibility (time burden, paperwork detail, etc), perceived invalidity of the measures for their specific patient population, and difficulty interpreting scores. These barriers are consistent with 2 of the 3 barriers (logistical concerns and appropriateness of measures) reported by Abrahamson.[20] There was strong consensus regarding frustration with time burdens for the completion of the measures and this was voiced clearly and strongly by clinicians who indicated feeling somewhat oppressed by "paperwork" demands. Instruments such as the CBCL and CAFAS were described as overly detailed and cumbersome. In addition, many clinicians reported concerns that the measures lacked cross-cultural sensitivity and that their clients' limited literacy might limit the validity of the scores. The majority of the clinicians reported receiving negative feedback from parents and youth about the administration of these instruments. However, this study did not directly assess family members' perceptions of the burdens of completing such forms.

The clinicians in this study clearly, and virtually unanimously, reported that the outcome measures were not clinically useful. However, many proceeded to give examples later in the interview of how the administration of the measures had been useful, practical, and informative. They made an implicit, yet very clear, distinction between the use (or lack thereof) of the scored feedback and the clinical process of giving the forms to youth and parents. This may, on a small scale, reflect a perceived or implicit dichotomy between the "imposed" empirical science and the experiential process of practice. The following quote dramatically expresses this dichotomy: *"I tend to look at processes and then, in order to talk about outcomes stuff, I take off one hat and then put on another, which requires me to decrease my intuitiveness, increase my intellect, and shut off my heart."*

Given the investment of resources in the administration, collection, and scoring of the performance outcome measures, it was disappointing to find that the vast majority of clinicians did not use the scored profiles that were provided to them. Many indicated that the scores were too difficult to understand, others simply felt that the scores were not a valid reflection of the client's functioning, or that the scores provided no new information. Commonly cited reasons for perceived invalidity included language and/or cross-cultural differences in interpretation, and inappropriateness for specific patient population (generally defined as "not right for my clients"). As suggested by Beutler, one of the essential challenges in mental health research and practice is the need to increase the "acceptability" of outcome measures for clinicians.[27]

Many of the clinicians reported difficulty interpreting the scored profiles for the performance outcome measures. All had attended training sessions and had received written guidelines for interpretation, but none reported using these reference materials to aid in score interpretation. Given the variability in training across and within disciplines, there is likely great variability in knowledge about outcome measures across mental health clinicians. Even those who reportedly knew how to interpret the scores expressed skepticism that the data would be of any clinical use. Most clinicians believe that outcome monitoring is extraneous to clinical practice.[28] Despite all the policy and research emphasis on outcome data, there has been minimal attention to the interpretation and use of such data in standard practice.[7,29] Recent developments in patient-focused outcome research have begun to address these challenges (see special section of *Journal of Consulting and Clinical Psychology*[29]). These promising efforts are designed to improve the feasibility, timeliness, and clinical utility of outcome monitoring. However, improved clinician training addressing misconceptions about outcome monitoring and demonstrating clinical utility of outcome assessment is needed.

The clinicians' suggestions for improved utility and feasibility of outcome evaluation include incorporation of briefer measures with simpler score interpretation. Many suggested narrative, as opposed to quantitative, summaries of scores. In addition, they suggested different outcome constructs to assess. The most common of these were client history data, such as history of maltreatment, learning disabilities, and previous mental health service use. These results are generally consistent with Bickman and colleagues' finding that clinicians valued client history data, such as maltreatment experience, most highly.[3] This is noteworthy in that most of the suggested variables are not outcome variables, but rather, they are immutable, historical experiences or characteristics. This suggests that clinicians are requesting data based on initial assessment needs, as opposed to ongoing evaluation of treatment progress.

One of the surprising findings of the study was that, despite all of the public and policy attention to increased accountability for providers, most clinicians reported that they did *not* feel pressured to document their effectiveness. Many expressed doubt that the outcome data would, or could, ever be used at the administrative level to evaluate individual clinicians. Others' statements reflected perceived "immunity" to empirical evaluation, suggesting that clients' perceptions of effectiveness were far more important and relevant than any empirical assessment. In fact, the majority of the clinicians' comments reflected confidence in their effectiveness, and this may be related to their lack of perceived accountability pressure. It is possible that this represents some selection bias in the sample (ie, "insecure" clinicians may have been less likely to participate in interviews).

Limitations

The limitations of this study include a less than optimum response rate of 43% of the eligible participants. However, this response rate is similar to other published survey studies of mental health professionals.[3,4] Very few nonparticipants actively refused to participate, although some "passively refused" by not responding to repeated calls. The extent to which the participants' attitudes and experiences are representative of all clinicians in this system, or clinicians in other locales and other types of non-public service systems, is not known. However, the participants are generally representative of the population of clinicians in this service system by gender, discipline distribution, years of experience, and age. While the sample is too small to definitively test for group differences in attitudes (eg, by discipline or years of experience), the exploratory analyses did not reveal any differences of large magnitude across these groups.

Implications for Behavioral Health Services

In this era emphasizing the delivery of evidence-based practice, it is critically important, and perhaps sobering, to examine the attitudes and experiences of clinicians in community-based practice settings. Although many clinicians were supportive of efforts to empirically evaluate the effectiveness of their services, an equal number expressed ideological opposition, or at least strong skepticism, regarding such efforts. Virtually all of the clinicians were frustrated by feasibility barriers of administering outcome measures, such as time burdens and clients' literacy challenges. Finally, although all of the participants had received scored profiles reporting their clients' performance on standardized measures, none reported using these results in treatment planning or monitoring.

These findings offer some constructive recommendations for successful implementation of outcome monitoring protocols. Pragmatic issues regarding feasibility of measure administration are obviously of great importance. Validity concerns such as cross-cultural and cross-clinical setting sensitivity should also be addressed. In addition, implementation and adoption of any new intervention is likely to be more successful and most relevant to the clinical context if all participants are involved in the implementation process from the start. Clinicians in this study expressed frustration that their input was never sought prior to the mandate being enforced. Input from clinicians with experience using outcome measurement is essential to identify and promote the clinical utility of the process.

The "ideological" barriers expressed by some clinicians regarding outcome evaluation may be most difficult to address, but are critically important to address for successful adoption of empirically supported intervention and assessment techniques in community-based practice settings. Methods to address clinicians' attitudes may need to come from related fields such as social cognition and marketing science.

Although outcome assessment protocols are intended to bring more empirical rigor to practice, we are concerned that some mandated efforts could further widen the gap between the science and practice of children's mental health services if clinicians' frustrations are not addressed and clinicians are not involved in the development and implementation of outcome monitoring protocols. In addition, the field must offer evidence of how the use of standardized outcome measurement benefits clients and families. It is somewhat ironic that the utility and benefit of outcome measurement is based largely on "intuitive sense and hope" when clinicians are being asked to shift away from reliance on intuition and hope.[7]

Acknowledgment

This study was funded by grant # MH 50144 from NIMH.

References

1. Koch JR, Lewis A, McCall D. A multistakeholder-driven model for developing an outcome management system. *The Journal of Behavioral Health Services & Research.* 1998;25:151–162.
2. Cesare-Murphy M, McMahill C, Schyve P. Joint commission evaluation of behavioral health care organizations. *Evaluation Review: Measuring Quality in Mental Health Services.* 1997;21:322–329.
3. Bickman L, Rosof J, Salzer MS, et al. What information do clinicians value for monitoring adolescent client progress? *Professional Psychology: Research and Practice.* 2000;31:70–74.
4. Addis ME, Krasnow AD. A national survey of practicing psychologists' attitudes toward psychotherapy treatment manuals. *Journal of Consulting and Clinical Psychology.* 2000;68:331–339.
5. Lueger RJ, Howard KI, Martinovich Z, et al. Assessing treatment progress of individual patients using expected treatment response models. *Journal of Consulting and Clinical Psychology.* 2001;69:150–158.
6. Burnam MA. Measuring outcomes of care for substance use and mental disorders. In: Steinwachs DM, Flynn LM, Norquist GS, et al, eds. *New Directions for Mental Health Services, No. 71: Using Client Outcomes Information to Improve Mental Health and Substance Abuse Treatment.* San Francisco, Calif: Jossey-Bass; 1996:3–18.
7. Smith RG. State of the science of mental health and substance abuse patient outcomes assessment. In: Steinwachs DM, Flynn LM, Norquist GS, et al, eds. *New Directions for Mental Health Services, No. 71: Using Client Outcomes Information to Improve Mental Health and Substance Abuse Treatment.* San Francisco, Calif: Jossey-Bass; 1996:59–67.
8. Goebel L. A peer review of feedback method of promoting compliance with preventive care guidelines in a resident ambulatory care clinic. *Joint Commission Journal on Quality Improvement.* 1997;23:196–202.
9. Dawes RM. *House of Cards: Psychology and Psychotherapy Built on Myth.* New York: Free Press; 1994.
10. Garb HN. *Studying the Clinician: Judgment Research and Psychological Assessment.* Washington, DC: American Psychological Association; 1998.
11. Meehl PE. *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence.* Minneapolis, Minn: University of Minnesota Press; 1954.
12. California Mental Health Planning Council (CMHPC). *Adult Performance Outcome Study: Wave 1 to Wave 2.* Sacramento, Calif: CMHPC; 1997.
13. Achenbach TM. *Manual for the Child Behavior Checklist/4–18 and 1991 Profile.* Burlington, VT: University of Vermont Department of Psychiatry; 1991.
14. Achenbach TM. *Manual for the Youth Self-Report and 1991 Profile.* Burlington: University of Vermont, Department of Psychiatry; 1991.
15. Hodges K. Child and Adolescent Functional Assessment Scale (CAFAS). In: Maruish ME, ed. *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment.* 2nd ed. Mahwah, NJ: Erlbaum; 1999:631–664.
16. Attkisson CC, Zwick R. The Client Satisfaction Questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. *Evaluation and Program Planning.* 1982;5:233–237.
17. Breda CS. Methodological issues in evaluating mental health outcomes of a children's mental health managed care demonstration. *The Journal Of Mental Health Administration.* 1996;23:40–50.
18. Hoagwood K, Burns BJ, Kiser L, et al. Evidence-based practice in child and adolescent mental health services. *Psychiatric Services.* 2001;52:1179–1189.
19. National Advisory Mental Health Council (NAMHC). *Blueprint for Change: Research on Child and Adolescent Mental Health. A Report by the National Advisory Mental Health Council's Workgroup on Child and Adolescent Mental Health Intervention Development and Deployment.* Bethesda, Md: National Institutes of Health/National Institute of Mental Health; 2001.
20. Abrahamson DJ. Outcomes, guidelines, and manuals: on leading horses to water. *Clinical Psychology: Science and Practice.* 1999;6:467–471.
21. Addis ME, Wade WA, Hatgis C. Barriers to dissemination of evidence-based practices: addressing practitioners' concerns about manual-based psychotherapies. *Clinical Psychology: Science and Practice.* 1999;6:430–441.
22. Aarons GA. Clinician personality and attitudes toward adoption of empirically supported treatments. Presented at the 15th International Conference on Mental Health Services Research. Evidence in Mental Health Services Research: What Types, How Much, and Then What? National Institute of Mental Health; April 2002; Washington, DC.
23. Hoagwood K, Jensen PS, Petti T, et al. Outcomes of mental health care for children and adolescents, I: A comprehensive conceptual model. *Journal of the American Academy of Child and Adolescent Psychiatry.* 1996;35:1055–1062.
24. QSR Nudist Vivo 1.0 [computer software]. Melbourne, Australia: Qualitative Solutions & Research Pty Ltd; 1999.
25. Beutler LE. David and Goliath: when empirical and clinical standards of practice meet. *American Psychologist.* 2000;55:997–1007.
26. Achenbach TM. *Manual for the CBCL/4–18 and Profile.* Burlington: University of Vermont Department of Psychiatry; 1991.
27. Beutler LE. Comparisons among quality assurance systems: from outcome assessment to clinical utility. *Journal of Consulting and Clinical Psychology.* 2001;69:197–204.
28. Persons JB. Conducting effectiveness studies in the context of evidence-based clinical practice. *Clinical Psychology: Science and Practice.* 2001;8:168–172.
29. Lambert MJ, Hansen NB, Finch AE. Patient-focused research: using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology.* 2001;69:159–172.