

Use of the Child and Adolescent Functional Assessment Scale (CAFAS) as an Outcome Measure in Clinical Settings

Kay Hodges, Ph.D.
Maria M. Wong, Ph.D.
Mark Latessa, M.S.W.

Abstract

This article discusses how the Child and Adolescent Functional Assessment Scale (CAFAS) can be used as an outcome measure in clinical settings. Outcome data from two clinical samples are provided: a small community mental health center located in Michigan and a large referred sample from the Fort Bragg Evaluation Project. Outcome indicators for assessing change over time included overall level of dysfunction, percentage of respondents with severe impairment, mean total score, mean scores for individual CAFAS subscales, and change in total score at the client level. Implications of the findings were discussed from several perspectives: improving services to individual clients, developing databases at the local level that can be used for the agency's continuing self-scrutiny, and pooling databases across sites that can be used to study broader issues within a managed care environment.

The Child and Adolescent Functional Assessment Scale (CAFAS)¹ is used as an outcome measure by publicly and privately funded entities. Many states currently use the CAFAS to determine eligibility for state-managed programs or for measuring performance-based outcomes.²⁻⁵ Individual agencies use the CAFAS to actively manage cases because it helps clarify treatment goals and track progress toward those goals over time. In addition, the CAFAS has been included as an outcome measure in two large studies: the Fort Bragg Evaluation Project (FBEP)⁶ and the evaluation of the demonstration grants funded by the Center for Mental Health Services.^{7,8}

In today's environment, use of measures is often mandated by oversight authorities or funding sources for the purposes of assessing outcome and determining client eligibility for specific levels of services. The individual agency is left to determine how the information can be used to benefit both the clients and the agency in general. It is critical that these measurements be clinically useful at the individual client level. Otherwise, it is likely that the instrumentation will not be applied reliably and, thus, will be unable to detect change over time.

In this article, information is presented on how to integrate the CAFAS into day-to-day clinical activities. Suggestions are also provided on how the information generated by the CAFAS can be used for the agency's internal evaluation. Data gathered by a small community mental health center (CMHC) in Michigan are used as a case example. The data were gathered as part of an informal pilot

Address correspondence to Kay Hodges, Ph.D., Department of Psychology, Applied Research Unit, Eastern Michigan University, 102 King Hall, Ypsilanti, MI 48197.

Maria M. Wong, Ph.D., is a postdoctoral fellow, Surrey Research Center, the Institute for Social Research, University of Michigan, Ann Arbor.

Mark Latessa, M.S.W., is a research and quality assurance specialist, Quality Improvement Services, Livingston County Community Mental Health Services.

project that preceded a statewide evaluation effort. Indicators that can be used to describe the youth served by the agency and to describe changes in client status over time are illustrated. In addition, the agency's data are compared to a large clinical sample collected in the FBEP for the purpose of external comparison. In the discussion, lessons learned are shared about how to maximize the accuracy and usefulness of the outcome data for internal evaluation.

Integration of the CAFAS in Clinical Care and Decision Making

In this section, the CAFAS, its psychometric properties, and ways in which the CAFAS can be integrated into routine clinical care are briefly described. A more detailed discussion is presented by Hodges.^{9,10}

The CAFAS comprises a set of written items describing behavior that are organized into domains of functioning. Within each domain, behaviors are grouped into levels of impairment: severe, moderate, mild, and no or minimal impairment. For each scale, the rater determines the level of impairment that best describes the youth's most severe level of dysfunction during the time specified by the user (e.g., last month, last 3 months). The scores assigned to each of the categories are as follows: 30 for severe, 20 for moderate, 10 for mild, and 0 for minimal or no impairment.

The CAFAS is rated by a staff member based on information collected as part of the typical clinical services. The actual rating usually takes about 10 minutes. A structured 30-minute interview has been developed to obtain the information needed to rate the CAFAS,^{11,12} although it is not necessary to use it. The 1989 version of the CAFAS, which was used by the CMHC and FBEP, consists of five scales for the youth and two for the caregiver. The scales used in this case study are as follows: Role Performance (i.e., how effectively the youth fulfills societal roles in school, home, and community), Behavior toward Self and Others (i.e., appropriateness of the youth's daily behavior), Moods/Emotions (i.e., modulation of the youth's emotional life), Thinking (i.e., ability of the youth to use rational thought processes), and Substance Use (i.e., youth's substance use and the extent to which it is inappropriate or disruptive).^{*} The total score refers to the sum of the five scales assessing the youth, with a range from 0 to 150. A higher score reflects greater impairment.

Satisfactory interrater reliability, test-retest reliability over a short time period, and internal consistency have been reported elsewhere.^{11,13} Concurrent and predictive validity of the CAFAS was generated by the FBEP, in which youths referred for mental health services were assessed at intake, at 6 months, and at 12 months. High discriminant validity was demonstrated; inpatients (i.e., psychiatric inpatient, residential treatment center) scored higher on the CAFAS than youths in alternative care (i.e., alternative care to traditional residential, including home-based services, day treatment, specialized foster care, and group home), who in turn scored higher than youths in outpatient care. This finding was observed at intake and at 6 months and 12 months postintake.¹³ As is desirable for a measure of outcome, the CAFAS proved to be sensitive to change. There was a significant main effect for time, with a reduction in CAFAS scores observed across time.¹³ The CAFAS has also been useful in predicting service utilization and costs at 6 months and 12 months postintake. Higher impairment was significantly related to more restrictive care, higher cost, more bed days, and more days of services. In addition, when compared to other measures and to clinical diagnosis, the CAFAS score was the strongest predictor of subsequent service utilization and cost at 6 months and 12 months postintake.¹⁴

The CAFAS generates clinically useful information at three levels: individual item endorsement, level of impairment on each of the individual scales, and summary score indicating overall

^{*} The current version of the CAFAS¹ has eight youth scales (School Role Performance, Home Role Performance, Community Role Performance, Behavior toward Others, Mood, Self-Harmful Behavior, Substance Use, and Thinking) and includes a list of strengths and/or goals appropriate to each scale.^{9,10}

dysfunction. The items endorsed are descriptions (e.g., expelled from school) that can be useful in developing goals on the youth's treatment plan. In the scoring summary, critical items indicating risk behaviors toward others or the self are brought to the rater's attention (e.g., fire setting, aggression, suicidal risk). The impairment level for each scale (i.e., severe, moderate, mild, and no or minimal) is determined by the items endorsed. Specific profile patterns or high endorsements on specific scales may trigger an evaluative consultation or suggest a particular treatment protocol. The profile provides a format for organizing discussions with the youth, his or her caregivers, or other professionals.

A total score is generated by summing the youth scales. There are no cutoff scores for the CAFAS. However, a general framework for putting the CAFAS total score into context, referred to as Overall Level of Dysfunction, appears to be useful for laypersons.⁹ For the 1989 version, the levels are as follows: 0 to 10 = none or minimal dysfunction, 20 to 30 = mild impairment (e.g., traditional outpatient services may suffice), 40 to 60 = moderate impairment (e.g., youths may be manageable with typical outpatient care or may need services beyond weekly outpatient visits), 70 to 80 = marked impairment (e.g., treatment that is more intensive than typical outpatient services would likely be warranted), 90 or higher = severe impairment (e.g., at minimum, an intensive intervention program would be indicated).

Case Study: Results for a CMHC

Procedures and Sample

A CMHC asked its staff to complete the CAFAS as part of a self-evaluation project. No extra funding or time was allotted to the clinicians, and since it was not required, there was no monitoring to determine compliance. Clients were to be rated on the CAFAS at their initial visit, after evaluation (if applicable), quarterly thereafter, and at discharge.

The sample presented in this article includes all cases for which there were at least two CAFAS ratings during a 15-month period. If there were more than two CAFAS administrations, the first and last CAFAS were used. For the purposes of this article, these two CAFAS administrations are referred to as the "intake" and the "last" CAFAS. As a result, the "last CAFAS" did not correspond to either a fixed time period or to the CAFAS done at discharge. However, the most frequent length of treatment was 6 months, and two-thirds of the cases only had two CAFAS ratings. The sample consisted of 179 youths, ranging in age from 5 to 17 years. The mean age was 11.15 years ($SD = 3.47$), with the breakdown of age categories as follows: 5 to 7 years old, 18.6%; 8 to 10, 26.6%; 11 to 13, 23.7%; and 14 and older, 31.1%. The sample was 66.5% boys and 33.5% girls. The only exclusions were to be emergency circumstances in which insufficient information to rate the CAFAS was available. Because data were not obtained on cases for which the CAFAS was not done, sample bias was unable to be examined.

Severity of Youths at Intake

The severity of the impairment of youths seen at intake can be summarized in three ways: (1) the mean for the total score; (2) Overall Level of Dysfunction, as indicated by CAFAS total score; and the (3) number of individual scales on which the youth was rated as severely impaired.

Mean, median, and mode for total score. The mean for the CAFAS total score was 60.78 ($SD = 25.01$). The scores ranged from 10 to 130, with a median of 60 and a mode of 50. The distribution of total scores was positively skewed (skewness = .55, $SE = .18$; kurtosis = $-.11$, $SE = .36$), indicating a longer tail at the higher end of the impairment scale. Almost 70% of the sample scored between 40 and 80, inclusively. The typical client was clearly impaired and would likely involve more management than the traditional "once a week, seen in the office only" treatment protocol.

Overall level of dysfunction. Another way of evaluating the sample at intake is to examine the frequency of youths whose total score at intake fell into each of the five categories of general functioning described above. Of the 179 youths, less than 1% ($n = 1$) scored in the lowest category (0-10), indicating no treatment likely needed. Only 13% ($n = 23$) of their cases would definitely be expected to be treated on a nonintensive, outpatient basis (scores of 20-30). Half of the sample ($n = 90$) had scores ranging from 40 to 60. Depending on the situation, youths at this level could potentially be treated on numerous points along a continuum of restrictiveness of care. More than one-third of the clients fell into the two severest categories. For the 19% ($n = 34$) with scores of 70 and 80, treatment that is more intensive than typical outpatient services would likely be needed. The 17% ($n = 31$) of youths scoring in the 90 and above range are very impaired. Their CAFAS total scores were more than one standard deviation above the average youth being seen in the clinic. Residential care may be justified if there are few resources available for managing the youth's behavior. On the other hand, managing these youths in nonresidential care, with no serious or long-lasting negative side effects on the youth, family, or community, would likely indicate considerable resources and strengths within the youth's environment. In this case study, it was known that no youths were in inpatient settings, in residential treatment centers, or placed outside the community, which reflected well on the agency.

Frequency of individual scales rated at the severe impairment level. For each client, the number of individual CAFAS scales on which the youth was rated as severe impairment (i.e., received a score of 30) was summed. The frequency of cases were as follows: no severe ratings ($n = 99$), one ($n = 43$), two ($n = 24$), three ($n = 12$), and four ($n = 1$). Thus, 45% of the clients were rated as severely impaired on one or more scales. One-quarter of the clients were severely impaired on only one scale, and the remaining 20% were impaired on two or more scales.

Change over Time

The indicators used to describe youth at intake can be used to assess change over time, comparing the first and last CAFAS rated.

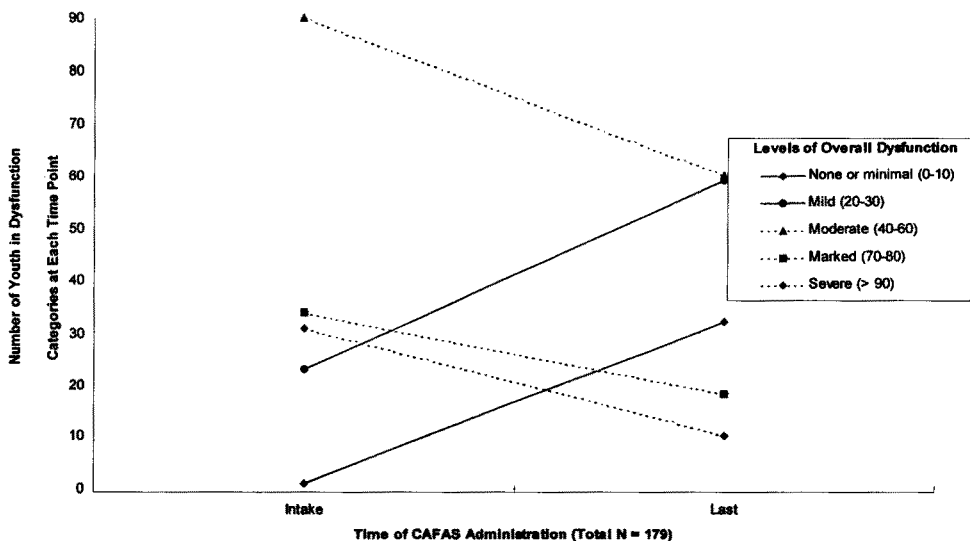
Mean, median, and mode for total score. Change over time was assessed with a paired t -test that compared the first and last CAFAS. The test was significant, $t(178) = 10.79, p < .0001$. The average score at the last CAFAS was 38.60 ($SD = 25.74$), compared to 60.78 ($SD = 25.01$) at intake. The reduction in the CAFAS total score was almost one standard deviation. The effect size was .86, which is above the mean observed for one-group pre-post designs based on a meta-analysis reported by Lipsey and Wilson.¹⁵ The median was reduced from 60 to 30, and the mode from 50 to 20. A score of 20 could be obtained by scoring mild on two scales, a condition that most parents probably would not consider serious enough to refer the youth for treatment. Thus, the modal score for the last CAFAS, which represents the most frequent score, would indicate a subclinical status.

Change in mean scores for individual scales. Using paired t -tests, the intake and last scores were compared for each of the five individual scales. Each of the t -tests was significant, except for the Substance Use scale. The results were as follows: Role Performance: $t(178) = 8.86, p < .0001$; Thinking: $t(178) = 4.53, p < .0001$; Behavior toward Self and Others: $t(178) = 8.31, p < .0001$; and Moods/Emotions: $t(178) = 9.16, p < .0001$. For each scale, there was a significant reduction in impairment as a function of time. The lack of significance for the Substance Use scale appears to reflect a floor effect in that there were very low impairment scores even at intake.

Overall level of dysfunction. If youths improved, an increase in the number of youths rated in the lowest levels of overall impairment (i.e., 0-10 and 20-30 groups) would be observed. For the remaining levels, all of which reflect a clinical level of impairment (i.e., the 40-60, 70-80, and 90 and higher groups), the number of youths should decrease. Figure 1 illustrates the change over time for

Figure 1

Change in Dysfunction Level from Intake to Last CAFAS in the CMHC Sample



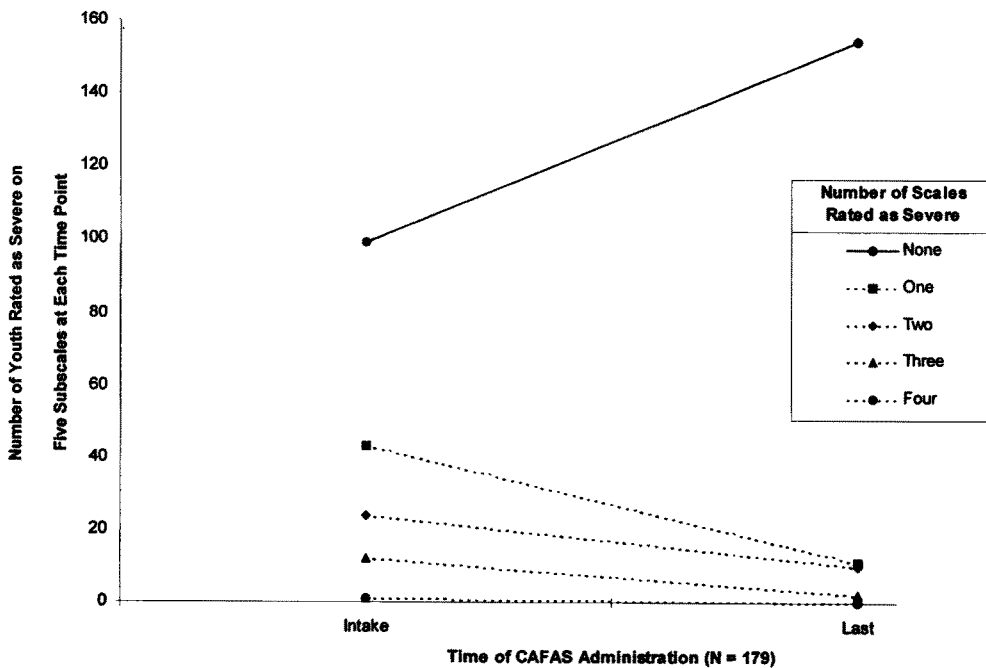
NOTE: CAFAS = Child and Adolescent Functional Assessment Scale; CMHC = community mental health center.

each of the five levels of overall dysfunction. The solid lines show how the number of youths in the two lowest categories increased from first to last assessment, and the nonsolid lines show that the number of youths scoring in the higher levels of impairment at intake decreased from first to last CAFAS. The percentage of youths without impairment (0-10 scores) went from less than 1% at intake to 18% at last CAFAS, and the percentage of youths with mild impairment went from 13% at intake to 33% at last CAFAS. More than half of the cases had either no impairment or a low level of impairment (i.e., score of 30 or less) on their last CAFAS. At the other end of the continuum, the percentage of youths in the most severe category (i.e., scoring 90 or higher) was reduced by two-thirds (i.e., 17.3% to 5.6%), and in the next most severe category (i.e., 70-80), the percentage of youths was reduced by one-half (from 19.0% to 10.1%). This indicator provides a means of pinpointing cases for further scrutiny. Supervising clinical staff would likely want to review the 10 cases whose overall level of dysfunction was the highest level on their last CAFAS.

Frequency of individual scales rated at the severe impairment level. Figure 2 shows the change over time of the number of scales that were rated as severely impaired. If youths get better over time, the number of youths having no scales rated as severe should increase, while the youths with one, two, or three scales rated as severe should decrease. In Figure 2, the solid line represents youths who had no scales rated as impaired and the nonsolid lines represent the other groups. A large majority of youths (87%) had no severe ratings of impairment on their last CAFAS. At the other extreme, only two youths had pervasive impairment, with three scales impaired. These two youths already would have been identified for further scrutiny in the above analysis because their total score would have been at least 90.

Change in total score at the client level. Another way to examine change is to determine whether each youth's total score increased, decreased, or stayed the same. "Stayed about the same" was

Figure 2
 Change in Number of Youth Rated as Severely Impaired
 on the Five CAFAS Subscales in CMHC Samples



NOTE: CAFAS = Child and Adolescent Functional Assessment Scale; CMHC = community mental health center.

defined as receiving the same score or a score within 10 points of the intake score. This definition would allow for some error in measurement given that the total score is measured in 10-point increments. Thus, a score that had increased was at least 20 points higher than the intake score, and a score was considered to have decreased if it was at least 20 points lower than the intake score. The majority of youths (59%) had lower impairment scores on their last CAFAS. About one-third (35%) scored in the same range. In all, 6% ($n = 11$) of the youths had higher impairment scores on their last CAFAS. Statistical testing could be conducted to test whether change in a particular direction (i.e., increased or decreased) was significant.

Record review of cases that appeared to deteriorate over time. The medical records for these 11 youths were reviewed. In this case study, there was no available Management Information System (MIS) that could be linked easily to the outcome data, so a qualitative review of the records was undertaken. The demographic characteristics of the youths were representative of the clinic, so there was no evidence that these youths disproportionately represented any subgroup of clients. For two cases, it appeared that the first CAFAS score may have been inaccurate in the direction of underestimating impairment. The clinic had a practice of rating the CAFAS at the first contact with the family (i.e., referred to as screening) and at the second contact (i.e., referred to as evaluation). Depending on the circumstances of the specific case and the flow of clients at the commencement of services, some clients were rated only once and others twice. Because of this practice, what appeared to be an obviously erroneous score given to two youths at screening could be identified. For 3 of the 11 cases, the clients had improved by the time the case was closed, which occurred after the ending date for the

study. Two of the cases represented unsuccessful outcomes, despite considerable effort having been expended in both cases. The extensive clinical work included multiagency collaboration with intensive home-based services. Review of the clinical notes suggested that perhaps these two families felt unable to keep the youths in their home, despite the good intentions of everyone involved. For the remaining four cases, the family prematurely withdrew from treatment. These observations were very useful to the agency for its own self-corrective actions.

Summary

Overall, the findings from this case study appeared to be very favorable. However, for them to be considered meaningful, do they need to be compared to the performance of comparable agencies serving similar youth? For the purpose of self-evaluation within a clinical setting, the answer is no. The target goal for each youth would be a low total CAFAS score, ranging from a score of 0 (i.e., no impairment) to perhaps 30 (i.e., low to moderate impairment), provided that no individual scale was rated as severe. Certainly, a postintervention total CAFAS score of 30 or lower would be desirable for every youth, although perhaps not realistic for all. This goal also makes common sense from the perspective of the content of the CAFAS items at the different levels of severity for each individual scale. For example, a layperson would likely agree that a youth with occasional problems obeying school rules (i.e., rated as mild impairment on the School scale) is much better off than a youth who has been expelled (e.g., from the severe level) or has such persistent problems that school authority figures know the youth (i.e., from moderate level). As such, the CAFAS has a standard in terms of desirable clinical outcome, and comparative data are not essential.

However, to the extent that each youth cannot achieve a level of no impairment or that some situations are difficult to change, comparable data from similar agencies serving similar clients would be desirable. Given the lack of comparative data at the time of this study, existing research data were used to provide a ballpark figure of what might be a reasonable benchmark. The analysis that was conducted for the CMHC was replicated with data from the FBEP. These benchmark data were helpful when the CMHC administrative staff informally presented information to their board members.

Comparison to a Large Referred Sample

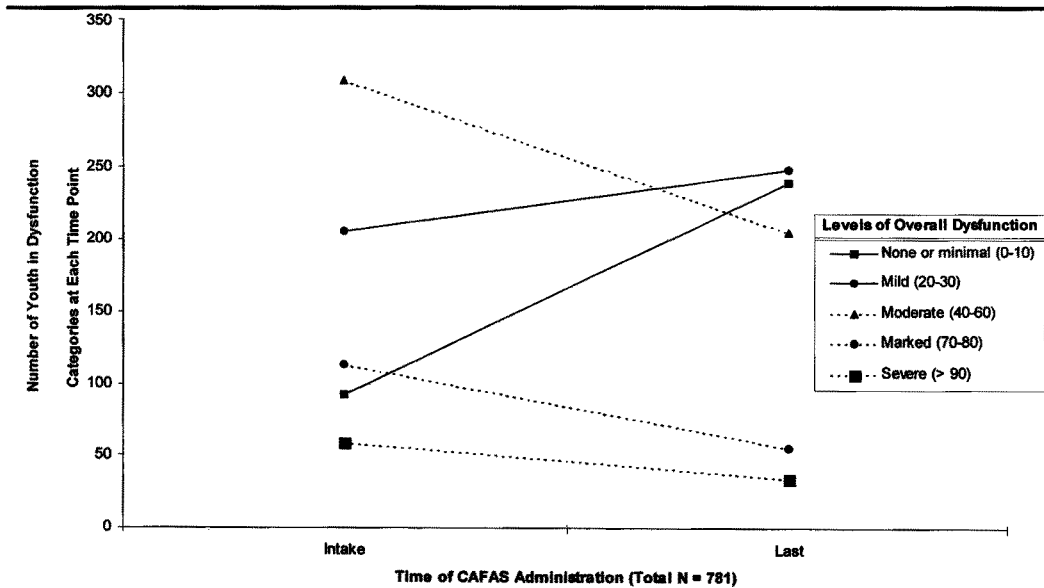
The first and second waves of the FBEP were chosen for comparison. The second wave, at 6 months, was chosen as opposed to the third wave, at 12 months, because more than two-thirds of the CMHC cases were closer to 6 months than to 12 months. A 6-month benchmark was thought to be the most useful comparison, since longer episodes of treatment under managed care will likely be the exception rather than the rule.

No statistical analyses were conducted comparing the two samples because the studies differ in many ways, making such analyses inappropriate. Examples of these differences include the timing of the second CAFAS rating (i.e., last vs. 6 months postintake), characteristics and training of the raters (i.e., clinicians vs. trained lay raters), economic incentives, and portal of entry procedures.^{16,17} More detailed information on the sample and methods used in this study can be found in Breda,¹⁶ Friedman,¹⁸ and Hodges and Wong.¹³

Sample and Procedure

The 781 youths were evaluated at intake and 6-month postintake in the FBEP. The sample consisted of all youths who were referred for mental health services and agreed to participate in the study. They were from one of three army bases: Fort Bragg, North Carolina; Fort Stewart, Georgia; or Fort Campbell, Kentucky. Thus, the data from three sites were collapsed. It is worth noting that because the purpose of the FBEP was to evaluate the services at Fort Bragg by comparing them to two comparison sites (Fort Stewart and Fort Campbell), published articles have emphasized the lack

Figure 3
 Change in Dysfunction Level from Intake
 to 6-Month CAFAS in the FBEP Sample



NOTE: CAFAS = Child and Adolescent Functional Assessment Scale; FBEP = Fort Bragg Evaluation Project.

of differences in outcome between the Fort Bragg and the comparison sites.^{6,16} However, there was a statistically significant reduction in impairment observed for both Fort Bragg and the comparison sites.⁶

Comparison on Severity of Youth at Intake

The mean for the CAFAS total score at intake was 45.65 ($SD = 26.47$). The scores ranged from 0 to 140, with a median of 40 and a mode of 50. The percentages of cases scoring in the overall dysfunction categories for the CAFAS total score were as follows: none or minimal impairment (0-10), 11.9%; mild (20-30), 26.4%; moderate (40-60), 39.6%; marked (70-80), 14.6%; and severe (≥ 90), 7.6%. The percentage of cases that were rated as severe on the CAFAS individual subscales at intake were as follows: none rated as severe, 87.2%; one severe rating, 9.7%; two severe ratings, 2.7%; three severe ratings, 0.4%; and four severe ratings, 0.0%.

Comparison on Change over Time

There was significant reduction in total CAFAS score from intake ($M = 45.65$, $SD = 26.47$) to 6 months postintake ($M = 31.39$, $SD = 26.03$), $t(780) = 14.33$, $p < .0001$. The effect size for change from intake to 6 months was .51. For the interval from intake to 12 months, the effect size was .67 ($t[616] = 16.67$, $p < .0001$), and for intake to 18 months, .78 ($t[372] = 15.00$, $p < .0001$). Paired t -tests were conducted for each of the individual CAFAS scales comparing intake to 6 months. Significant differences were observed on each of the five scales in the direction of less impairment at 6 months.

Change in overall level of dysfunction, as indicated by the impairment categories for the total CAFAS score, is depicted in Figure 3. The same pattern as was identified for the CMHC sample was observed. The number of youths scoring between 0 and 10 and between 20 and 30 increased, while the number of youths in each of the remaining categories (i.e., scores of 40 and higher) decreased. In

the CMHC and FBEP samples, at the last (or 6-month) time point, 85% to 89% of the youths scored 60 or lower.

At 6 months, the data on percentages of youth rated as severely impaired on the five CAFAS subscales were as follows: none, 89.6%; one, 7.3%; two, 2.7%; three, 0.4%; and four, 0.0%. In both samples, at the last (or 6-month) CAFAS, about 85% to 87% of the youth were not severely impaired on any subscale. However, the data for FBEP were not as impressive as the data for the CMHC sample, because 87% of the FBEP sample had no severe impairments at intake.

Change in the total score at the client level for the FBEP was as follows: 45.7% had a lower impairment score 6 months later, compared to the intake total score; 42.1% scored within 10 points of the intake score when they were evaluated at 6 months; and 12.2% had a higher impairment score at 6 months, compared to intake.

Summary

Compared to this benchmark, the CMHC sample at intake was rated as more severe on each of the three indicators. This could be due to any number of factors, such as accuracy of the CAFAS ratings, influences that attracted less impaired youth for referral (i.e., no co-pay, all youth having to use one portal in the FBEP), and/or bias in sampling or attrition.

As for change over time, on each of the indicators, the CMHC performed as well or better than the FBEP. In part, this would be expected because the CMHC was rated as more impaired at intake. In presenting their statistics, CMHC staff could state that their 6% deterioration rate was less than the 12% observed in the FBEP, in which resources for treatment were extremely generous.

An important caveat is that no assumption about the causative factors for the reduction of impairment can be made for the case study or for the FBEP. There was no control group because no clients were denied treatment; nor were clients intentionally delayed in receiving treatment so that a waiting-list control group could be engineered. As a consequence, no unequivocal reason, such as effective treatment, can be attributed to change observed in the clients across time. However, this is the typical scenario in applied settings. In fact, it is hard to imagine that most funding sources for mental health services would want to pay for rigorous research or approve of conducting research on clients that includes denying or delaying treatment.

Discussion

This pilot study was initiated by the CMHC because the child and adolescent staff wanted to do a better job of assessing their effectiveness. To their credit, this was the first step in a process that has evolved into a statewide initiative. The result is a partnership in which local service providers, state administrators, and university researchers cooperate to generate an empirical basis that can hopefully influence policy decisions. The aim is to conduct mental health services research along the lines described by Speer and Newman.¹⁹ The process by which this collaboration unfolded will be described briefly, followed by a discussion of specific suggestions regarding the use of the CAFAS as an outcome indicator.

Service providers wanted some help with assessing outcome, which was soon to be mandated at the state level, and with developing a data-based feedback loop with the aim of improving their services. As a result of consultation from the Research and Training Center of the University of Florida, various stakeholders around the state came to understand that for outcome data to be useful, a broader array of information needed to be collected (e.g., client characteristics, interventions offered).²⁰ This resulted in the providers of child services deciding to gather more data than required by the state and to pool information across sites. State administrators were being asked to generate a set of guidelines for determining levels of care within a managed care environment. To their credit, they wanted input from various stakeholders and welcomed the opportunity to empirically study any proposed guidelines. The availability of an automated version of the CAFAS made data collection

and aggregation across sites feasible. The CAFAS program permitted computer scoring of the CAFAS as well as collection of other information about the youth, the youth's environment, and services offered. A state university had opened an applied research program and was eager to assist in aggregating and analyzing the data for the state administrators and the service providers. A pooled data set was generated for the purpose of developing an empirical basis for policy decisions. This statewide initiative is still ongoing; however, some suggestions based on experience with the case study are described.

Implications for Behavioral Health Services

Maximizing Credibility of Ratings

In applied settings, it is seldom possible to have persons independent of the clinical services complete outcome measures. In fact, a strong case can be made that the treating clinician is the best person to make judgments about the client's growth. The clinician or case manager should be the professional who knows the most about the client. However, the credibility of the clinician's ratings may be questioned because of the obvious motive to show improvement in the client. Even so, from a practical and ethical perspective, the clinician may be the best judge. So, how can the veracity and credibility of the ratings be maximized given these constraints?

Probably the most important factor in reducing bias is the fact that each rating on the CAFAS must be supported by endorsement of at least one behavioral description of the youth. The CAFAS instructions stipulate this, and the form is designed to make justification easy to do. This is in contrast to global measures that do not require explicit justification of ratings.^{21,22} To maintain the integrity of the measure, it is critical that CAFAS ratings on each scale be supported by at least one specific item endorsement. Thus, requiring only scale scores for each scale would be unwise. This would endanger the accuracy of the rating. It is only human nature to allocate the least time and attention to a task that does not require justification or does not "really count" as part of the medical record. Clinicians are accustomed to the notion that notes in the medical records should be accurate; to do otherwise can be viewed as a fraudulent act. Thus, requiring the rater to specify which CAFAS items support the rating and incorporating the CAFAS form or profile into the medical record help reduce inaccurate ratings.

In addition, assessing interrater reliability and ensuring that all raters are satisfactorily trained will reduce rating error not due to demand characteristics. For this reason, the CAFAS training materials are designed so that they can be easily used and can be used to actually test the trainee's interrater reliability.^{9,23} In fact, the CAFAS Self-Training Manual²³ is designed so that it can be done independently, although staff often enjoy group training because it provides an opportunity to discuss common concerns. Even very experienced raters need to establish interrater reliability because the important issue is that all raters follow the same rules. For example, is using extremely profane words an act of aggression toward an authority figure? No measure can be sensitive to true changes in performance unless it is reliably rated.

Other practices that should help reduce inadvertent and intentional rater bias include auditing and use of independent raters. From an administrative perspective, this issue is partly addressed via audits. For example, one state hired an auditing firm to examine a sample of the medical records to determine if there was support for the CAFAS ratings and for the measure being used for adults. Feedback was given to the individual sites so that they could improve their documentation. Another approach is to have independent CAFAS ratings done by trained raters. In fact, with a small sample of the CMHC clients, graduate students trained in the CAFAS conducted the CAFAS interview with a parent on the telephone.¹² The interview takes about 30 minutes and obtains all of the information needed to rate the CAFAS. The reason for the interview is that the independent rater does not have

the same means of obtaining information as does the treating clinician. For example, a managed care company used the telephone interview approach to obtain CAFAS ratings on a sample of its clients whose services were being funded through the department of social services. The state required that providers report on outcome of services provided to their clients. This same approach can be used to conduct follow-up outcome studies for costly residential services in which the youth's behavior is constrained by external controls. Assessing outcome in the postdischarge time period after return to the community provides the most credible evidence.

Clinical Usefulness of the Outcome Data

Clients were to be evaluated at intake, every 3 months thereafter, and at discharge. The evaluations were to be used by frontline staff in making ongoing decisions about the client as treatment progressed. It was assumed that the CAFAS profile would be shared with caregivers, who were collaborators in shaping treatment interventions. Outcome assessment was conceptualized as a dynamic process that actually influenced treatment decisions, not as a measure to be done as pre- and post-events. As such, the measure becomes an actively used tool in the case management process. Assessment becomes part of the treatment in the hands of the clinician.

In the statewide initiative that evolved, this concept was preserved. The CAFAS computer program produced a client report each time a CAFAS evaluation was done, with a graph comparing the youth's intake and current CAFAS scores. Unless an outcome measure can be used for ongoing monitoring of the client's progress, it will be seen as an added paperwork burden with little redeeming value for the client or the frontline staff. Having the outcome measurement incorporated into the clinical routine in a meaningful way contributes to "buy in" on the part of frontline staff.

Quality Assurance

The CAFAS indicators that track change over time can identify youths who are not progressing well or who are functioning poorly even after some intervention. After identifying the types of cases that do not progress well, these cases can be monitored prospectively. The outcome indicators can be used to evaluate treatment protocols developed for specific client profiles. In fact, data can be pooled over sites to identify the strengths of various programs. Rather than compete, programs can cooperate around their areas of strength.

If the CAFAS indicators are linked to utilization data, then severity of the youth's impairment at intake can be compared to services offered, services received, and cost of services. Examining this relationship should indicate whether level of care received matched the apparent level of need. Exceptions to this expected pattern can be examined to see if customary care was offered. Cases that would warrant further scrutiny would be low-impairment cases with restrictive, out-of-community, or costly placements and high-impairment cases that use few resources or few liaisons with other service agencies.

This information can be used to develop realistic goals about treatment options for various types of client profiles. How much can a youth benefit and what is the typical course of change? When do the costs outweigh the marginal gains? These are difficult issues that deserve empirical investigation.¹⁹

Acknowledgments

The authors would like to thank the professional, support, and administrative staff of the Child and Adolescent Services of Livingston County Community Mental Health Services (LCCMH). In particular, we want to express appreciation to the members of the board of the LCCMH, Angus M. Miller, executive director; Larry Newberg, associate director of clinical affairs; and Sherry Whalen, formerly program director of child and adolescent services. We also want to acknowledge the

support of the Michigan Department of Community Health, especially Sherida K. Falvay, director of the Office of Children's Services, and Jim Wotring, director of Programs for Children with Emotional Disturbance.

References

1. Hodges K: *Child and Adolescent Functional Assessment Scale*. Ypsilanti: Eastern Michigan University, Department of Psychology, 1989, 1994, 1997.
2. Behar LB, Stelle L: Criteria for accessing child mental health and substance abuse services in North Carolina. In: Liberton CJ, Kutash K, Friedman RM (Eds.): *Proceedings of the Annual Research Conference: A System of Care for Children's Mental Health: Expanding the Research Base*. Tampa, FL: Research & Training Center for Children's Mental Health, 1997, pp. 262-264.
3. Irvin E, Hersch P: Proposed eligibility criteria and procedures for enrollment in Department of Mental Health continuing care. In: Liberton CJ, Kutash K, Friedman RM (Eds.): *Proceedings of the Annual Research Conference: A System of Care for Children's Mental Health: Expanding the Research Base*. Tampa, FL: Research & Training Center for Children's Mental Health, 1997, pp. 264-267.
4. Lemoine RL, Speier T, Ellzey S, et al.: Using the Child and Adolescent Functional Assessment Scale (CAFAS) to establish level-of-need for Medicaid managed care services. In: Liberton CJ, Kutash K, Friedman RM (Eds.): *Proceedings of the Annual Research Conference: A System of Care for Children's Mental Health: Expanding the Research Base*. Tampa, FL: Research & Training Center for Children's Mental Health, 1997, pp. 267-270.
5. Hodges K, Rosenblatt A, Irvin E, et al.: *Developing an Empirical Base for Managed Care in Public Mental Health Systems*. Symposium presented at the 10th Annual Research Conference: A System of Care for Children's Mental Health: Expanding the Research Base, Tampa, FL, February 1997.
6. Lambert WE, Guthrie PR: Clinical outcomes of a children's mental health managed care demonstration. *Journal of Mental Health Administration* 1996; 23(1):51-68.
7. Doucette-Gates A, Liao Q, Sondheimer A, et al.: *CMHS Evaluation: Model Changing—Metaphors and Meaning*. Symposium presented at the 10th Annual Research Conference: A System of Care for Children's Mental Health: Expanding the Research Base, Tampa, FL, February 1997.
8. Hodges K, Latessa M, Pernice F, et al.: *Practical Issues in Using the CAFAS for Clinical and Administrative Outcome*. Symposium presented at the 10th Annual Research Conference: A System of Care for Children's Mental Health: Expanding the Research Base, Tampa, FL, February 1997.
9. Hodges K: *CAFAS Manual for Training Coordinators, Clinical Administrators and Data Managers*. Ypsilanti: Eastern Michigan University, Department of Psychology, 1997.
10. Hodges K: Child and Adolescent Functional Assessment Scale. In: Maruish ME (Ed.): *The Use of Psychological Testing for Treatment Planning and Outcome Assessment*, in press. Liberton C, Kutash K, Friedman R: The 10th Annual Research Conference Proceedings, A System of Care for Children's Mental Health: Expanding the Research Base, 1998.
11. Hodges K: Psychometric Study of a Telephone Interview for the CAFAS Using an Expanded Version of the Scale. Paper presented at the 8th Annual Research Conference: A System of Care for Children's Mental Health: Expanding the Research Base, Tampa, FL, February 1995.
12. Pernice F, Gust J, Hodges K: *A Structured Interview for Collecting Objective Outcome Data in Clinical Settings*. Symposium presented at the 10th Annual Research Conference: A System of Care for Children's Mental Health: Expanding the Research Base, Tampa, FL, February 1997.
13. Hodges K, Wong MM: Psychometric characteristics of a multidimensional measure to assess impairment: The Child and Adolescent Functional Assessment Scale. *Journal of Child and Family Studies* 1996; 5:445-467.
14. Hodges K, Wong MM: Use of the Child and Adolescent Functional Assessment Scale to predict service utilization and cost. *Journal of Mental Health Administration* 1997; 24:278-290.
15. Lipsey MW, Wilson DB: The efficacy of psychological, educational, and behavioral treatment. *American Psychologist* 1993; 48:1181-1199.
16. Breda CS: Methodological issues in evaluating mental health outcomes of a children's mental health managed care demonstration. *Journal of Mental Health Administration* 1996; 23(1):40-50.
17. Foster EM, Summerfelt WT, Saunders RC: The costs of mental health services under the Fort Bragg demonstration. *Journal of Mental Health Administration* 1996; 23(1):92-106.
18. Friedman RM: The Fort Bragg Study: What can we conclude? *Journal of Child & Family Studies* 1996; 5:161-168.
19. Speer DC, Newman FL: Mental health services outcome evaluation. *Clinical Psychology Science and Practice* 1996; 3:105-129.
20. Hernandez M, Hodges S, Macbeth G, et al.: *Michigan Outcome Identification Project*. Tampa: University of South Florida, Florida Mental Health Institute, Department of Child and Family Studies, 1996.
21. Hodges K: *Measures for Assessing Impairment in Children and Adolescents*. Paper prepared for the U.S. Center for Mental Health Services. Rockville, MD: Department of Health and Human Services, Substance Abuse, and Mental Health Services Administration, 1994.
22. Hodges K, Gust J: Measures of impairment for children and adolescents. *Journal of Mental Health Administration* 1995; 22:403-413.
23. Hodges K: *The Child and Adolescent Functional Assessment Scale Self Training Manual*. Ypsilanti: Eastern Michigan University, Department of Psychology, 1994.