Prediction of Outcome from the Dartmouth Assertive Community Treatment Fidelity Scale

By Gary R. Bond, PhD, and Michelle P. Salyers, PhD

FOCUS POINTS
• Assertive community treatment (ACT) is an evidence-based practice used to monitor fidelity of implementation are critical.
• The Dartmouth ACT fidelity scale is a useful tool for monitoring ACT programs.

ABSTRACT

Background: Assertive community treatment (ACT) is an intensive and comprehensive treatment for clients with severe mental illness (SMI) who do not readily benefit from clinic-based services. Monitoring the implementation of such programs is critical, because better-implemented programs have been found to be effective in improving client outcomes.

Objective: We tested the hypothesis that fidelity to the ACT model would be positively correlated with improved client outcomes, as measured by reduction in psychiatric hospital use.

Methods: A scale measuring fidelity of program implementation, the Dartmouth ACT Scale, was examined in 10 newly formed ACT teams. Using the team as the unit of measure, the mean reduction in state hospital days for a 1-year period before and after program admission was calculated. Mean effect size in reduction in hospital days was used as the outcome measure in a correlational design.

Results: Pre/post comparisons showed a 43% reduction in hospital days for 317 clients (t=8.61, P<.001). The Pearson correlation between DACTS fidelity and reduction of state hospital days was .49, P=.08, one-tailed.

Conclusion: Several possible reasons are offered for why the study hypothesis was not confirmed. However, even if predictive validity of the Dartmouth ACT Scale is limited, it continues to be a useful tool for program monitoring and for providing corrective feedback.

INTRODUCTION

Assertive community treatment (ACT) is an intensive and comprehensive treatment for clients with severe mental illness (SMI) who do not readily benefit from clinic-based services. It was developed in the 1970s as a community alternative to psychiatric hospitalization. In the ACT model, a multidisciplinary team meets daily to plan interventions for clients who are assisted in home and community settings with medications, housing, finances, and a range of everyday issues and concerns. Low client: staff ratios allow for frequent contact with clients. ACT teams provide services directly rather than referring clients to other agencies. Admission criteria for ACT teams typically include frequent prior hospitalizations or long-term hospitalizations, co-occurring substance use disorders, homelessness, and/or involvement with the legal system.

ACT has been widely recognized as an evidence-based practice (EBP). National dissemination of ACT has been strongly recommended in government reports, by expert consensus panels, and through the advocacy of the National Alliance for the Mentally Ill. ACT is one of the most extensively researched models of community care for people with SMI. The evidence for the effectiveness of ACT is quite consistent across numerous reviews.

Although shown to enhance independent living, improve quality of life, and decrease psychiatric symptoms, its impact is strongest in reducing hospital use among clients at risk for psychiatric hospita...
tion. Because of costs of repeated hospitalizations and limitations in conventional mental health services in treating such clients, ACT has appealed to state mental health planners. Consequently, numerous states have implemented statewide initiatives to provide ACT services.

Despite the fact that the core ingredients of ACT are well-articulated and widely understood, implementation of ACT programs has been variable. This lack of fidelity (ie, adherence to program standards) in ACT programs has led to the need for methods to monitor implementation. State mental health authorities, practitioners, researchers, and consumer groups all have sought out tools for assessing adequacy of implementation of program models. Explicit quantitative monitoring of program fidelity is increasingly recognized as an important quality assurance technique. Over the last decade there have been many different independent efforts to develop ACT fidelity scales. The informal use of such scales has been widespread, despite meager information on their reliability and validity.

Three researcher-developed ACT fidelity scales have been reported in the literature. The Index of Fidelity to ACT (IF-ACT) was devised by McGrew and colleagues. IF-ACT assessed objective features of a program, such as the inclusion of a nurse on the team, frequency of team meetings, and frequency of contacts with clients in the community. It consisted of 17 items, divided into three subscales: Staffing, Organization, and Services. The items were retrospectively coded on 18 ACT programs in completed studies. IF-ACT proved to have adequate internal consistency. Moreover, it had good predictive validity, significantly correlating with a program-level measure of reduction in days in psychiatric hospitals ($r=.06, P<.001$).

Around the same time, Teague and colleagues independently developed a 13-item scale to assess fidelity of implementation of an ACT approach adapted for clients with SMI and co-occurring substance use disorders. The content of their scale overlapped with IF-ACT, although the scoring system was different. Teague and colleagues demonstrated that the ACT programs generally scored significantly higher on their fidelity scale than the control condition (standard case management). In a subsequent analysis, McHugo and colleagues found that among the clients in their sample, those in high-fidelity ACT programs had higher rates of retention in treatment, greater remission from substance use disorders, and fewer hospital admissions, than those in low-fidelity programs.

Among the ACT fidelity scales reported in the literature, the most widely used is the Dartmouth ACT Scale (DACTS). It was developed to address some limitations in scaling, explicitness, and comprehensiveness of the earlier fidelity scales. The majority of criteria and anchors in the new instrument were adapted from variables used as indicators for the 13 items reported by Teague and colleagues. Additional variables were designed on the basis of results reported by McGrew and colleagues. A primary goal was to have a measure that could discriminate well-implemented ACT programs from other types of case management services and at the same time provide an accessible tool for training and self-evaluation within programs. The DACTS was piloted in 50 case management programs, representing four distinct types of service models: ACT; intensive case management provided by the Veterans Administration; outreach programs for people who were homeless and mentally ill; and traditional case management. The DACTS discriminated across the four types of case management, consistent with predicted order of similarity to ACT. These findings were corroborated in a second analysis using an expanded database. Other studies have suggested that the DACTS has adequate internal consistency, acceptable interrater reliability, and sensitivity to change over time.

Although three studies have found an association between fidelity to ACT and improved client outcomes, no published papers have examined this relationship for the DACTS, despite its extensive use. This article describes an exploratory study of the DACTS...
as a pragmatic management tool to monitor newly formed ACT teams in a statewide initiative. Our question was whether ACT fidelity would be associated with better client outcomes, as measures by the DACT.

METHODS

Sample

In 1994, the Illinois state mental health authority launched a statewide initiative to develop ACT programs for frequent users of state psychiatric hospitals. This initiative was inspired by the unprecedented success of the Chicago-based Thresholds Bridge programs29-32 in reducing state hospital beds in a series of program demonstrations. The Thresholds Bridge was an early adaptation of the ACT model for this target client group.

Ten service providers were selected to receive special state funding in the initiative, mostly located in communities having the highest rates of heavy users of state hospitals. (Two additional rural sites providing scaled-down ACT services and located in areas with modest hospitalization rates were excluded from this analysis). Each new ACT team was given a roster of eligible clients from their area who met one of the following criteria regarding use of state psychiatric hospitals: Three or more admissions in the last year; five or more admissions in the last 2 years; or >180 or more days hospitalized in the previous year.

Each ACT team received funding for five full-time case managers, a full-time supervisor, and a part-time psychiatrist. Program guidelines included the following elements: client-to-staff ratio not to exceed 15:1; programs expected to make aggressive and persistent attempts to engage, enroll, and serve clients; team meetings at least twice a week; shared caseloads, with clients seen on a regular basis by different ACT workers; psychiatrist as a member of ACT team with adequate time commitment; contacts with clients principally in community settings; an average of two contacts per week with each client; 24-hour coverage by the ACT team to permit rapid response to emergencies; and assistance to clients with money management.

Each site received additional funding for residential services and substance abuse treatment.

Fidelity

The DACTS consists of 28 items, each rated on a behaviorally anchored 5-point scale (1=not implemented; 5=fully implemented). In the current study, three items were excluded: two added after the initial psychometric study25 and one (dual disorder treatment groups), because it was contrary to the state’s ACT guidelines. We used the mean score for the total 25-item scale as the measure of ACT fidelity. According to norms from established ACT programs,25 a DACTS score of ≥4.0 is considered moderately high fidelity.

Outcome

We measured reduction of days in state psychiatric hospitals, obtained from the state’s management information database. In the current study, we used the site as the unit of measure, calculating the mean reduction in state hospital days for a 1-year period before and after program admission, and converting this value to an effect size, essentially a standardized measure of change taking into account the variation within each site.15 An effect size <.33 is considered “small,” between 0.33 and 0.55 is considered “medium,” and >.55 is considered “large.”34

Procedure

This study was approved by the IUPUI and Clarian institutional review boards. Client outcome data were collected as part of a 12-month program evaluation of this statewide ACT project. Staff from ACT teams completed a battery of instruments measuring client outcomes at baseline and at 6 and 12 months after program admission. Only hospital data were examined in this report.

Fidelity assessments were conducted by two advanced doctoral students in clinical psychology with extensive experience in several prior statewide ACT demonstration projects. They made three fidelity site visits to each ACT program over a 2-year period. During these site visits, assessors attended team meetings, accompanied case managers on home visits, reviewed progress notes, and informally interviewed staff about the structure and functioning of the ACT team. During the final site visit, one assessor conducted a structured interview with the team leader, which was used as the basis for the DACTS ratings in this report.

FINDINGS

Client Characteristics

Twelve-month pre/post client data were obtained for 317 clients, representing 75.4% of the 420 clients who had been ACT clients for ≥12 months at the time of the final report. Another 21 clients were closed within 12 months of program admission. Demographically, in the total sample (N=317), 65.4% were male; the mean age was 37.8 (SD=9.4); 33.4% were Caucasian,
64.9% were African American and 1.6% were other; 5.2% were currently married; and 50.3% had at least a high school diploma or equivalent; 80.6% were receiving supplemental security income or social security disability insurance; and 15.8% had worked for pay in the past 6 months. The majority had a diagnosis of schizophrenia spectrum disorder (70.5%); 16.4% had a diagnosis of affective disorder; and 13.1% some other psychiatric disorder. Over their lifetime, clients averaged 16.5 psychiatric hospitalizations (SD=15.4) and 38.4 months (SD=53.7) in state psychiatric hospitals. Staff identified 40.5% of clients with a co-occurring substance abuse diagnosis, while 33.2% of clients reported substance abuse problems. Staff reported that, in the 6 months prior to intake, 20.9% had had legal problems and 6.4% had been in jail; 23.3% had periods of time in the last year without any income. The sample averaged 26.6 days homeless (SD=75.5) in the year prior to ACT services.

**Hospitalization Outcomes**

Compared with a mean of 101.8 days (SD=113.3) hospitalized during the year prior to ACT admission, clients averaged 57.8 days (SD=94.8) during the year after admission, \( t=8.61, P<.001 \), a 43% reduction. At the site level, effect sizes for reduction in hospital days ranged from 0.17–0.88. In three sites the effect size was small, in four sites it was medium, and in three sites it was large.

**Fidelity-Outcome Relationship**

According to the DACTS, sites varied in their implementation from marginal to high fidelity, with DACTS scores ranging from 3.6–4.2 (mean=3.9, SD=0.2). As a whole, the Illinois ACT sample resembled the levels of implementation found in a sample of 14 established ACT teams (mean=4.0, SD=0.4), as reported in the original psychometric study of the DACTS.\(^{25}\) The Pearson correlation between DACTS fidelity and reduction of state hospital days was a moderately large but nonsignificant correlation in the predicted direction (\( r=.49, P=.08 \), one-tailed). As shown in the Figure, the distribution of DACTS scores divided sites into two subgroups: three teams with DACTS scores below 3.8, all below the median on outcome; and seven sites with DACTS scores >3.9, five (71%) of which were above the median on outcome. Put another way, of the three sites with small effect sizes for reducing hospital days, 1 (33%) had above-average fidelity. Of the four with medium effect sizes, three (75%) had above-average fidelity. All three sites with large effect sizes had above-average fidelity.

**DISCUSSION**

Fidelity scales for evidence-based practices are in their infancy.\(^{17,20}\) Much work is needed in defining these scales and showing their psychometric adequacy. Studies demonstrating reliability for most fidelity scales, including the DACTS, are largely absent from the literature. With regard to validity, one important step is showing that fidelity scales discriminate between programs that exemplify the practice from those that do not. This has been well established for the DACTS. What has not been shown is whether the DACTS predicts better outcomes. To date, no studies of the predictive validity of the DACTS have been reported.

The study hypothesis, that fidelity would be significantly correlated with reduction in state hospital days, was not confirmed. One interpretation is that fidelity, as measured by the DACTS, does not have a strong influence on program outcomes. The DACTS has been criticized as being too focused on structural and organizational components of the program model, to the exclusion of clinical elements. The findings are consistent with this interpretation. Alternatively, two other factors likely influenced the results. First, the small sample size limited statistical power, by requiring a very large correlation in order to reach significance. (The same correlation in a sample of as few as 13 sites would have been statistically significant.) The second factor was restriction of range in fidelity scores among program sites. Even the lowest-fidelity ACT programs in the current study were substantially higher than the fidelity for traditional case management, or even typical intensive case management programs.\(^{26}\) One implication is that, in states that are systematically implementing ACT, or any other EBP, when there is systematic state planning and training, we can expect that most sites will attain a minimal level of fidelity. Reinforcing this point is a recent statewide ACT implementation in Indiana, in which 10 of 11 ACT teams achieved DACTS fidelity score exceeding 4.0, generally regarded as a minimum standard for adequate fidelity.\(^{28,35}\) We can contrast the current study with an earlier study\(^{22}\) that did find a significant correlation between ACT fidelity and hospital reduction. In the earlier study, guidelines and training were far less systematic than in the current project. As the field becomes more sophisticated in providing effective infrastructure support for implementing EBPs,\(^{21}\) we should expect only modest correlations between fidelity and outcome among programs actively seeking to implement the practice. In this context, modest correlations will not necessarily mean that fidelity is unimportant or that the scales are invalid, but instead could signal successful implementation strategies.
In contrast to the restriction in range for DACTS fidelity scores, there was no restriction in range for the outcome variable. The ACT teams varied widely in their effectiveness in reducing hospital use. By design, all the teams began with a sample of frequently hospitalized clients. How do we explain the fact that some achieved large effects while others achieved only small effects on hospital reduction? We can safely conclude for the current study that fidelity of implementation was not a sufficient condition for achieving expected outcomes for ACT. Leadership, the quality of the interventions, the clinical skills of the staff, and many other factors should be considered for a full understanding of the impact of ACT services on outcomes. Research needs to be devoted to understanding the relative influences of such factors.

While not statistically significant, the moderately large correlation found in the current study is consistent with the view that fidelity of implementation may have been a factor in reducing hospital use among clients with a history of frequent hospitalization. Replication studies are obviously needed, but in a meta-analysis, the positive correlation in the current study would contribute toward a conclusion that fidelity is a factor.

Fidelity scales are commonly used not only for research purposes but also for quality assurance purposes. In the current project, to facilitate fidelity of implementation, DACTS assessments were used in individual feedback sessions held with each site. Present from the community mental health center were the ACT team leader and the center director. Also present were state mental health authority representatives and the research team. Discussions centered on an individualized feedback report, which included a site-specific table displaying DACTS item scores and the site's total score, with comparative aggregated data from the statewide sample and comparative data from established ACT teams. The feedback sessions had several intended purposes. Providers could observe areas in which they attained program standards and compare their progress with norms for established programs. For sites achieving at least moderate implementation, the review reinforced their progress. At the same time, the feedback sessions allowed for each site to identify, in concrete and specific terms, areas in which the site fell short of full implementation. The use of the DACTS to provide structured feedback was well received by state officials and provider agencies.

The Figure suggests a further application, which combines fidelity and outcome data. This graphical representation offers a diagnostic tool to direct technical assistance efforts. The fact that two of the three sites with below-average fidelity also had negligible effects on hospital use shows the practical utility of the fidelity ratings and shows how fidelity and outcome assessment can be used together to reinforce the need for change. Trainers also may target programs low in fidelity and those with poor outcomes despite high fidelity, which may have implementation problems (eg, clinician inexperience) not measured by fidelity scales.

This study has several important limitations. First, as previously noted, the small sample obviously limited statistical power. Second, the methods used did not permit examination of interrater or test-retest reliability. Third, the 25% rate of missing client-level data limits the validity of the outcome measure. Fourth, only one outcome measure was chosen for examination. Hospital use was chosen because of its sensitivity to ACT interventions, based on prior ACT reviews. A more comprehensive study would have included multiple outcomes.

**CONCLUSION**

In the mental health services field, fidelity assessment is a new but increasingly popular aid to implementation, dissemination, quality assurance, and process research. Only within the last decade have fidelity scales begun to gain widespread acceptance as an essential tool in both research and practice. Accordingly, there has been growing recognition of a critical need for broader application and further study of fidelity measures, both in regard to specific program implementation and overall program quality. The use of fidelity scales as dependent measures in a national study examining the implementation of EBPs highlights the importance of research linking fidelity to client outcomes. Although replication in larger samples is needed, our findings support the practical utility of the DACTS. Of course, any practical measure of fidelity, even a widely used one such as the DACTS, represents only an approximation of the full range of key program elements that account for a program's success.

Routine use of fidelity scales provides an objective, structured way to give feedback about program development. Experience also suggests that this is an excellent method to diagnose program weaknesses while helping to clarify program strengths. Fidelity scales also provide a comparative framework for evaluating statewide implementation of an EBP and allow the evaluator to identify statewide trends and outliers. Another use for the DACTS is in clinical supervision to stimulate discussions of clinical processes viewed...
as critical to achievement of outcomes, including elements not incorporated on the DACTS. In Iowa, state trainers are using the DACTS in still another innovative way: ACT practitioners provide fidelity assessments for ACT teams in other centers, providing the foundation for the formation of a learning community (Nancy Williams, MD, oral communication, June 2004). The strategic use of repeated evaluations of programs using fidelity scales, either for an individual program or on a statewide level, is based on the general principle that whatever is attended to is more likely to be improved.

REFERENCES