Published in final edited form as:

J Behav Anal Health Sports Fit Med. 2009 ; 2(1): 43–54.

MONITORING FIDELITY IN THE ADOLESCENT COMMUNITY REINFORCEMENT APPROACH (A-CRA): THE TRAINING PROCESS FOR A-CRA RATERS

Bryan R. Garner, Ph.D.[Research Scientist],

Chestnut Health Systems, 448 Wylie Drive, Normal, IL 61761, (309) 451-7809

Brandi Barnes, M.S.[EBT Coordinator], and

Chestnut Health Systems, 448 Wylie Drive, Normal, IL 61761

Susan H. Godley, Rh.D.[Senior Research Scientist]

Chestnut Health Systems, 448 Wylie Drive, Normal, IL 61761

Bryan R. Garner: brgarner@chestnut.org; Brandi Barnes: bbarnes@chestnut.org; Susan H. Godley: sgodley@chestnut.org

Abstract

Ensuring evidence-based treatments are delivered with a high degree of fidelity is an important aspect of transporting these practices to community-based treatment providers. Just as training is critical for clinicians who plan to deliver the Adolescent Community Reinforcement Approach (A-CRA), a process of training and vetting individuals who assess sessions for fidelity also is critical. This article describes the training process for session raters who assess fidelity of A-CRA during the clinician training and certification process. A-CRA is currently being implemented in 32 independent community-based agencies as part of a large initiative funded by the Center for Substance Abuse Treatment.

Keywords

training; treatment fidelity; adolescent; treatment; A-CRA

The importance of treatment fidelity has been recognized for nearly three decades (Sechrest, West, Phillips, Redner, & Yeaton, 1979; Yeaton & Sechrest, 1981). Moreover, it remains an important issue for many of today's health services researchers and practitioners (Barber et al., 2006; Borrelli et al., 2005; Orwin, 2000; Waltz, Addis, Koerner, & Jacobsen, 1993), especially in light of the growing emphasis on the use of evidence-based treatment (EBT; Garner, in press; Institute of Medicine [IOM], 1998; Miller, Sorensen, Selzer, & Brigham, 2006; Weisner et al., 2004). Confidence in a study's internal validity is increased when there is good documentation that the interventions under study have been implemented as planned. Additionally, continuing to measure fidelity is important as interventions are transported to practice, given it is challenging to achieve the levels of intervention fidelity attained during randomized clinical trials. Indeed, Henggeler and his colleagues at the Medical University of South Carolina have demonstrated the importance of assessing the degree of implementation in studies examining the transportability of Multisystemic Therapy (MST; Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 1998). For example, Henggeler, Melton, Brondino, Scherer, and Hanley (1997) examined the role of

treatment implementation in MST outcomes across two public sector mental health sites. Results suggested the observed therapeutic effects of MST on outcomes were minimal when implementation variation was not taken into account. However, when implementation variation was statistically controlled, the effects were consistent with those found in prior clinical trials of MST. Similarly, a meta-analysis of outcome studies conducted by Curtis, Ronan, and Borduin (2004) found the average effect of MST was much larger in efficacy studies ($d = .81$) than in effectiveness studies ($d = .26$).

Despite its importance, ensuring that a treatment intervention is delivered as intended by model developers can be very challenging. One important consideration in making sure a treatment intervention is delivered with high fidelity is identifying the components of the treatment intervention to be assessed. The Adolescent Community Reinforcement Approach (A-CRA; Godley et al., 2001) requires a unique approach for fidelity assessment. While A-CRA is a manual-based EBT, clinicians are not expected to deliver treatment procedures in a prescribed order or deliver every A-CRA procedure during every session. Rather, A-CRA clinicians are expected to select from a menu of different A-CRA procedures and introduce appropriate procedures when clinically indicated. Another issue related to fidelity measurement is the design of the methods/procedures to be used to assess these treatment components. A recent review of methods for the training and monitoring of behavioral interventions in multi-site addictions research conducted by Baer and colleagues (2007) suggests there are not established standards for treatment fidelity measures. Rather, specific components assessed and the methods or procedures for assessing them are often related to the complexity of the intervention(s) of interest and/or the studies' primary objectives. However, wide-scale implementation of an EBT in community settings, in contrast to research studies, elevates the importance of developing efficient fidelity rating approaches.

Many fidelity instruments for various EBTs have been shown to demonstrate good psychometric properties (e.g., Barber, Mercer, Krakauer, & Calvo, 1996; Carroll et al., 2000; Henggeler & Bourdin, 1992; Hogue et al., 2008; Madson, Campbell, Barrett, Brondino, & Melchert, 2005; Moyers, Martin, Manuel, Hendrickson, & Miller, 2005; Pierson et al., 2007). For example, Barber and colleagues (1996) described the development of a fidelity instrument for individual drug counseling (IDC) for treatment of cocaine dependence. Using a seven-point scale, three raters assessed 43 items based on the main components described in the treatment manual. These measures were used to assess both 'adherence' (i.e., extent to which intervention components are delivered as prescribed in the treatment manual), and 'competence' (i.e., qualitative measure of the skillfulness in which intervention components are delivered). Rater training included a practice period during which raters met weekly to rate 72 tapes, followed by discussion about how they were using the scale. Raters then rated 62 new tapes in order to assess interrater reliability. Overall, intraclass correlation coefficients (ICCs) for adherence and competence ratings were acceptable and ranged from .70–.89 and .55–.85, respectively. In a similar study, Carroll and colleagues (2000) described the development and validation of the Yale Adherence and Competence Scale (YACS). Composed of 55 items generated from both treatment manuals and review of session videotapes, the YACS assesses elements both common to most behavioral interventions (i.e., assessment, general support, and goals of treatment) and specific to 12-step facilitation or cognitive-behavioral treatment. Using a five-point scale, five raters assessed adherence and competence on each of the 55 items for 19 randomly selected sessions. Overall, interrater reliability was very good for each of the six YACS subscales, with ICCs ranging from .80–.95 for adherence and .71–.98 for competence.

Not all fidelity scales developed for EBTs have been shown to have good interrater reliability. As part of a randomized clinical trial comparing Cognitive Behavioral Treatment (CBT) and Multidimensional Family Therapy (MDFT), Hogue et al. (2008) examined the

interrater reliability of the Therapist Behavior Rating Scale – Competence (TBRSC-C). Described as an observational measure of both adherence and competence in individual and family-based treatment, the TBRSC-C assesses general therapeutic goals (e.g., establishing a working relationship, drug use monitoring) as opposed to focusing on specific techniques. Using a seven-point scale, seven CBT raters and eight MDFT raters assessed each of the instrument measures for adherence and competence. In all, 437 sessions were rated using the TBRSC-C (245 MDFT sessions and 192 CBT sessions). Although interrater reliability was acceptable for each of the adherence ratings (ICCs ranged from .52 to .79 for MDFT and .56 to .83 for CBT), interrater reliability for competence ratings was generally poor (ICCs ranged from .15 to .55 for MDFT and .01 to .63 for CBT). Based upon variance component analysis, client effects accounted for more of the explained variance in adherence and competence scores than did therapist effects; however, the vast majority of variance was unexplained. It is possible that the largest portion of variance is attributable to differences between the individual raters (i.e., rater effects). Unfortunately, this was unable to be estimated since two raters were randomly assigned to each tape, as opposed to all raters or a consistent subset of raters.

The purpose of this article is to present an overview of the adolescent version of CRA, and then to describe a large-scale implementation of A-CRA supported by the federal government in over 32 sites, as well as the process used to promote fidelity to A-CRA. The challenge of developing a method of training raters and rating A-CRA sessions for an implementation of A-CRA for over 30 sites simultaneously was great. To date, there had been no publication that provided guidance about how to train the number of raters required for such a large-scale implementation of CRA. Additionally, we provide data regarding background characteristics of raters who have completed the process, summary data on the rater training process, lessons learned, and directions for future research.

A-CRA and the Assertive Adolescent and Family Treatment (AAFT) Initiative

A-CRA is an adaptation of the Community Reinforcement Approach (CRA; Hunt & Azrin, 1973) that was evaluated in the Cannabis Youth Treatment (CYT) study (Dennis et al., 2002, 2004). The latter was a large randomized clinical trial that compared different outpatient interventions for adolescents with substance use problems. Initial CRA studies were conducted primarily with adult samples (e.g., Azrin, Sisson, Meyers, & Godley, 1982; Higgins et al., 1991; Hunt & Azrin, 1973; Smith, Meyers, & Delaney, 1998), though there was one early adolescent study that used some similar procedures (Azrin, Donohue, Besalel, Kogan, & Acierno, 1994). The adolescent version of CRA, A-CRA, was adapted by including examples of dialogue addressing adolescent issues, the addition of parent/caregiver involvement in four treatment sessions, revision of a tool used to measure satisfaction with life areas, and examples of adolescent-specific goals and activities (Godley et al., 2001). The manual describes a menu of different procedures (e.g., Functional Analysis of Substance Use, Happiness Scale, Problem Solving) from which clinicians draw based on what an adolescent and his/her caregiver discuss in session. Analysis of the outcomes from the CYT study revealed that on average, adolescents in each of the five outpatient interventions significantly improved during the treatment phase. However, there was a statistical trend for the A-CRA condition to have the highest percentage of adolescents in recovery (i.e., no alcohol or other drug use, abuse, or dependence symptoms while living in the community) 12 months after intake (34%) compared with the other two conditions in its study arm. Additionally, A-CRA was found to be statistically more cost-effective per person in recovery 12-months after intake than the two other interventions in its study arm (Dennis et al., 2004).

Due to the demonstrated effectiveness and cost-effectiveness of A-CRA, the Substance Abuse and Mental Health Administration's Center for Substance Abuse Treatment (CSAT) awarded 15 grants to community-based organizations to implement the A-CRA as part of its Assertive Adolescent and Family Treatment Initiative (AAFT; RFA TI-06-007) in September 2006. Most sites implemented A-CRA as an outpatient intervention with durations of 12 to 14 weeks, followed by another 12 to 14 weeks of in-home Assertive Continuing Care (ACC) during which clinicians also used A-CRA procedures. Exactly one year later, 17 additional grants were awarded as part of a second AAFT cohort for a total of 32 sites. These community-based sites were located in various locales around the United States and collectively served diverse populations in multiple types of settings (e.g., urban, rural, outpatient, residential, home-based). In general, the purpose of these service grants was to help community-based treatment agencies address the large treatment gap for adolescents needing intervention for their alcohol and other drug abuse/dependence by providing training for clinicians and supervisors in A-CRA. To help ensure success of this large-scale dissemination and implementation initiative, CSAT not only provided up to \$300,000 per year for three years to each of the 32 grantees, but also provided free training and technical assistance via a training and technical assistance contract awarded to Chestnut Health Systems. Thus, the AAFT project represents one of the field's largest initiatives to date to replicate an adolescent-focused EBT in practice settings.

Clinician Certification Process and A-CRA Rater Training

Clinician certification process

In the AAFT project, A-CRA raters provide fidelity checks as part of clinician and supervisor certification processes, and then as part of ongoing random fidelity monitoring. The purpose of these dual certification processes is to assure there are 'markers' that the funder (CSAT) and the technical assistant contractor can use to assess the degree to which clinicians and supervisors are progressing in the implementation process, and to provide specific feedback on fidelity for each clinician. Furthermore, the process is instrumental in creating a system at the local site that facilitates sustainability of A-CRA once CSAT funding is completed. Since sites and A-CRA raters are located all over the United States, the certification process relies on a website tool that allows: (a) uploading of digitally recorded sessions; (b) streaming of session audio for review by raters; and (c) posting of numeric ratings and narrative comments for review by clinicians and supervisors, as well as by "coaches" in preparation for telephone coaching sessions.

The clinician and supervisor certification processes are different, and for purposes of this paper, we briefly describe the clinician certification process. Overall, there are 19 A-CRA procedures that consist of 77 components, including two overarching skill sets: (a) overall adherence to the model and (b) general clinical skills. There is an A-CRA checklist (see Smith, Lundy, & Gianini, 2007) that corresponds to each procedure and required components. The training team selected nine core procedures that clinicians are required to "pass" in order to receive certification as an A-CRA clinician. After A-CRA certification, clinicians are asked to continue recording their sessions, in part to attempt to demonstrate competency on the remaining ten procedures. In addition, randomly chosen digital session recordings (DSRs) are reviewed every other month after clinician certification, and clinicians and their supervisors are provided numeric and narrative feedback based on fidelity reviews of these DSRs.

A-CRA rater training

Initially, A-CRA raters were graduate students who were trained extensively by one of the model developers. A few clinical psychology doctoral students continue to be trained and

serve as A-CRA raters, but as the number of clinicians in the certification process grew, the training team made a decision to recruit and train raters from the clinicians or supervisors who work in community-based agencies like the clinicians whose sessions they rate. Currently, the majority of A-CRA raters are community-based clinicians and clinical supervisors who work on a contractual basis. The training team identified individuals who had performed well in the clinical and/or supervision certification process themselves, and approached them about participating in training and certification as an A-CRA rater. Thus, these individuals had the experience of recording their clinical sessions and having a rater evaluate them and provide feedback. In deciding whom to recruit, the team also has attempted to identify individuals who live in different areas of the country, and who represent and work with different cultural groups. A related consideration in selecting raters was obtaining a subset of individuals who could rate sessions conducted in languages other than English (e.g., Spanish, Vietnamese), since the AAFT sites serve very diverse populations.

Several of the steps for A-CRA rater training are the same as for clinical or supervisor certification. An individual must first attend a national A-CRA training in order to become familiar with the A-CRA procedures and general clinical approach. Prior to attending the training, attendees are asked to read the A-CRA manual and take an A-CRA knowledge test until they are able to achieve a score of 80% or better. The 3.5-day training provides an overview of both the research supporting the model and the A-CRA procedures, and allows multiple opportunities to role-play and receive feedback for different procedures. Since most individuals are certified clinicians prior to being asked to become an A-CRA rater, they have already attended training before they begin the rater training process. After gaining familiarity with the model from attending the training and for most, completing one of the A-CRA certification processes, rater trainees are required to participate in an orientation phone call with a project coordinator. During this call, the coordinator reviews general decision rating rules (e.g., when to leave blanks for incomplete procedures), provides a live web demo of the password-protected EBTx website, and gives instructions regarding how to submit practice session ratings.

After the orientation meeting, the rater trainees are assigned their first session. They continue to be assigned test sessions to rate until they achieve agreement ratings of at least 80% in comparison with exemplar ratings for two sets of ratings across six or more procedures. The two rating agreements computed are for procedures the rater heard attempted (Procedures Attempted) and for the quality of the attempt (Quality Rating). Recall that when clinicians use A-CRA, they draw from a menu of procedures; therefore, it is not expected that every A-CRA procedure would be heard or rated in each session. Since only one or two procedures typically are delivered during any given session, a Procedures Attempted category is needed to capture this rating. The Quality Rating agreement rate provides a percentage for the accuracy of the reviewer's ratings compared to those of the exemplar ratings. For each component of each procedure that is tracked, a 1–5 rating scale is used, whereby 1 = poor, 2 = needs improvement, 3 = satisfactory, 4 = very good, 5 = excellent. Ratings are considered a match if there is no more than a one point difference between the rater's and the exemplar rating. An exception to this rule occurs when the two ratings are a 2 and a 3. Once a component of a procedure is rated a 3, it is considered satisfactory, while a rating of 2 is still considered unsatisfactory (i.e., not passing). Thus, ratings of 2 and 3, while only a one-point difference, are not considered to be in agreement. As an example of meeting the 80% matching requirement, a rater would need to match at least five of the six components (83%) that comprise the Functional Analysis of Substance Using Behavior procedure. After every rating attempt, rater trainees check an Excel spreadsheet to compare their posted ratings to the exemplar ratings, and to see the calculated agreement rates.

There is a CRA/A-CRA coding manual that contains operational definitions of each rating possibility (1–5) for each component of the 19 A-CRA procedures (Smith et al., 2007). Using the Functional Analysis of Substance Using Behavior procedure as an example again, note that the six required components include: Gave rationale, started by asking for description of common episode/behavior, outlined triggers (external; internal), clarified the using behavior, outlined positive and negative consequences of the behavior, and gave examples of how the information would be used. The rating manual specifies that in order to receive the best Quality Rating (5) on the rationale component, the clinician must provide a rationale that has three of the following: A description of at least some of the information that would be collected and why, a more general overview of plans for the information collected, an explanation of the Functional Analysis chart, and an explanation of how the information will be used in treatment. The clinician also must present this information ‘concisely and coherently’ (Smith et al., 2007, p. 27). Rater trainees are not given the coding manual until after they have demonstrated acceptable agreement rates across three of the six required sessions. This practice is followed in an attempt to get them to first experiment in a trial-and-error manner with the coding process themselves. The concern was that new rater trainees would get caught up in providing the “correct number” from a manual for a rating, and in the process would lose sight of the big picture.

Currently, there are 40 standard practice sessions used for rater training. These sessions cover a variety of the A-CRA procedures and were conducted by over 18 different clinicians. They range from very poor to excellently-conducted sessions. Based on which procedures the rater trainee is able to rate with acceptable agreement, the project coordinator determines the next session to be assigned. If a procedure is passed with a consistency of 80% or better, a session with a different procedure is then assigned for review. The goal is to provide rater trainees the opportunity to evaluate a wide variety of the A-CRA procedures during the training process.

Being able to provide detailed narrative feedback is considered a critical dimension of the rater training and certification process, because this feedback tells clinicians what they did well and how they can improve their delivery of the intervention. Thus, the rater training process also includes shaping as to how the raters provide this feedback. The rating manual offers sample narrative comments that are appropriate when giving feedback on each procedure. Rater trainees receive a sample narrative feedback after rating a test session. After a rater has achieved the desired agreement rate on four of the six required sessions, they are then required to submit a narrative piece for evaluation in addition to the ratings. Raters are taught to provide this feedback in an “A-CRA manner”, which means that there is a heavy emphasis on providing positive feedback wherever possible, as well as specific comments to shape the desired clinical behaviors.

Summary Statistics on A-CRA Expert Raters and Their Training

Currently, the AAFT initiative is in the full implementation stage with 95 clinicians from all 32 AAFT sites uploading DSRs to the website. On average, approximately 86 DSRs are uploaded each week, with approximately 37 being requested for either certification or post-certification review. Thus, far more than 2,300 DSRs have been reviewed, with 1.4 hours required on average to complete a single DSR review. As most A-CRA raters have part- or full-time positions as clinicians, many only provide ratings on a part-time basis, which has required the recruitment of a large number of raters. Since the start of the AAFT initiative, a total of 31 individuals have been asked to consider completing the process to become an A-CRA expert rater. None of these individuals refused to agree to pursue rater training, but seven (23%) have not begun the process or completed it in a timely manner. A total of 24 individuals (77%) have completed the process of becoming A-CRA raters. Of these

individuals, 75% are Caucasian, 21% Hispanic, and 4% African American; 92% are female and have an average age of 33 years ($SD = 7.2$). Seventy-one percent have a master's degree, with the remaining 29% having a bachelor's degree. On average, raters have more than five years of substance abuse counseling experience and an average of two years of CRA/A-CRA experience. Of these 24 raters, 20 have achieved expert A-CRA rater designation, while four are still in the process. For the 20 individuals who have achieved the expert rater designation, approximately 10 attempts were needed on average before they were able to achieve the necessary agreement rate of 80% across six session ratings. Although the number of hours spent training was not logged, it took raters an average of 83 days from when they began rating DSRs until they had attained their expert rating status.

Lessons Learned

As is clear from the literature reviewed in the introduction, the measurement of fidelity is important when evaluating or implementing any EBT. In the CSAT project, the need to implement A-CRA in such a large number of community-based agencies simultaneously presented unique challenges. Several components were needed for the technical assistance model for this large project including: (a) standardization of an A-CRA rating measure; (b) completion of a ratings manual to increase the agreement of ratings across raters; (c) development of a training process that helped determine when a rater was ready to provide the ratings and narrative feedback that would be helpful to clinicians, supervisors, and A-CRA coaches; and (d) a therapist certification process, so that there was a means of monitoring how well sites were actually progressing in implementing the model faithfully.

A number of lessons were learned from this development process. One of these is the need for ongoing training with raters. The training can take many forms. For example, we allow therapists or supervisors to request re-review of their ratings or narrative comments if they feel that they are not valid. The review completed by a second rater is then compared with that conducted by the first rater. If discrepancies are found, this information is shared with the first rater. This process often provides valuable feedback to individual raters regarding their accuracy. Additionally, every clinician is asked to provide feedback once they have completed the certification process. Some of these e-mails are very specific about how rater comments could have been more helpful to a clinician in improving her/his A-CRA delivery. Coaches, who routinely use rater comments during bi-monthly coaching calls, also provide feedback to raters about how to improve the clinical utility of their ratings. Finally, there is a monthly call for all raters that is conducted by one of the model developers. Raters receive feedback during this call about specific rating issues that have arisen, and time is allotted for raters to raise and discuss new rating questions.

The methods of rating A-CRA sessions differs in some significant ways from fidelity ratings of other EBTs, since A-CRA is a procedure-based model and not all procedures are required to be used in every session. For example, if an adolescent does not want to work on his substance use during a session, but appears motivated to talk about problems with a teacher at school, it would be appropriate for the clinician to use A-CRA procedures related to either problem-solving or communication skills during that session and not discuss the adolescent's substance use directly. This highly flexible approach offers a number of therapeutic advantages, but has implications for fidelity ratings. Thus, while some rating approaches only require raters to review 10 or 20 minute segments of sessions, as is the case for motivational interviewing (Moyers et al., 2005; Pierson et al., 2007), A-CRA raters must listen to an entire session to allow the raters to determine whether the most appropriate A-CRA procedures were attempted, and then to rate how competently these procedures were delivered.

Also as previously described, since not all A-CRA procedures are expected to occur in a session, one component of the rating scale is the indication of whether or not a given procedure was used. We have found that such information can be useful in examining the mechanisms of change for A-CRA. For example, Garner et al. (2009) recently showed that A-CRA exposure, a dimension of treatment fidelity which represents a count of the number of different A-CRA procedures delivered to adolescents and their parents, was not only significantly related to treatment outcome, but fully mediated the relationship between treatment retention and outcome

Another unique aspect of the A-CRA fidelity process is the individualized approach to rater training. Specifically, whereas other EBTs train a group of raters with the goal of achieving an overall level of interrater reliability exceeding a certain threshold, A-CRA rater training is individualized. In other words, the goal for each A-CRA rater is to demonstrate the ability to consistently rate sessions and different procedures with a high level of agreement to a standardized set of ratings. Importantly, interrater reliability only indicates a group of raters have demonstrated the ability to give ratings in a consistent manner and may not relate to the accuracy of these ratings. For instance, if all raters always gave the highest possible rating, this would result in perfect interrater reliability, but would not be accurate. Thus, assuming exemplar ratings provided by individuals trained by model developers are indeed accurate, training individual raters to some standard suggests that fidelity ratings have at least face validity.

Directions for Future Research

There are several research questions our group would like to address in the future. Future research using rigorous designs is needed to identify the best and most cost-effective methods for training fidelity raters and for efficient ongoing monitoring of treatment fidelity in practice. As we have noted, approaches may differ for specific EBTs. For example, although A-CRA competence is rated on a 5-point scale similar to the rating scales used for other fidelity instruments, a unique aspect of the A-CRA rating process is that a rating of 3 represents the point at which a minimum level of competency is said to have been demonstrated. An important implication of this process is that the one-point difference between a 2 and 3 is qualitatively different from the one-point difference between a 3 and 4. Future research is needed to examine to what extent this distinction is clinically meaningful (i.e., related to treatment process and posttreatment outcomes). In addition to how well the respective treatment fidelity rating process relates to important treatment outcomes, future research needs to focus on how well those outcomes replicate those in the efficacy studies in which the intervention was originally evaluated. Fortunately, as 3-, 6-, and 12-month follow-up data will be collected on over 4,000 adolescents who participate in A-CRA as part of the CSAT project, we will be able to examine the relationship of A-CRA fidelity ratings to outcomes within the initiative, and conduct a comparison of the findings between the original implementation of A-CRA in the CYT effectiveness trial with its implementation within this initiative. Lastly, as part of our future research we plan to evaluate the interrater reliability across A-CRA raters and the relationship of ratings to a rating by one of the authors of the rating manual.

Acknowledgments

This work was supported by contract 270-2003-00006 with Center for Substance Abuse Treatment (CSAT), Substance Abuse and Mental Health Services Administration (SAMHSA), the National Institute on Drug Abuse (NIDA grant DA018183), and the National Institute on Alcohol and Alcoholism Abuse (NIAAA grant AA017625). The authors wish to thank Mark Godley, Robert J. Meyers, Randy Muck, Jutta Butler, and Kelli Wright for their work on this project. The authors also wish to acknowledge Kelli Wright for her assistance in preparing this

manuscript for submission. The opinions are those of the authors and do not represent official positions of the government.

References

- Azrin NH, Donohue B, Besalel VA, Kogan ES, Acierno R. Youth drug abuse treatment: A controlled outcome study. *Journal of Child and Adolescent Substance Abuse*. 1994; 3:1–16.
- Azrin NH, Sisson RW, Meyers R, Godley MD. Alcoholism treatment by disulfiram and community reinforcement therapy. *Journal of Behavior Therapy and Experimental Psychiatry*. 1982; 13:105–112. [PubMed: 7130406]
- Baer JS, Ball SA, Campbell BK, Miele GM, Schoener EP, Tracy K. Training and fidelity monitoring of behavioral interventions in multi-site addictions research. *Drug and Alcohol Dependence*. 2007; 87:107–118. [PubMed: 17023123]
- Barber JP, Gallop R, Crits-Christoph P, Frank A, Thase ME, Weiss RD, Gibbons MBC. The role of therapist adherence, therapist competence, and alliance in predicting outcome of individual drug counseling: Results from the National Institute Drug Abuse Collaborative Cocaine Treatment. *Psychotherapy Research*. 2006; 16:229–240.
- Barber JP, Mercer D, Krakauer I, Calvo N. Development of an adherence/competence rating scale for individual drug counseling. *Drug and Alcohol Dependence*. 1996; 43:125–132. [PubMed: 9023068]
- Borrelli B, Sepinwall D, Bellg AJ, Breger R, DeFrancesco C, Sharp DL, Ernst D, Czajkowski S, Levesque C, Ogedegbe G, Resnick B, Orwig D. A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology*. 2005; 73:852–860. [PubMed: 16287385]
- Carroll KM, Nich C, Sifry R, Frankforter T, Nuro KF, Ball SA, Fenton LR, Rounsaville BJ. A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*. 2000; 57:225–238. [PubMed: 10661673]
- Curtis NM, Ronan KR, Borduin CM. Multisystemic treatment: A meta-analysis of outcome studies. *Journal of Family Psychology*. 2004; 18:411–419. [PubMed: 15382965]
- Dennis ML, Godley SH, Diamond G, Tims FM, Babor T, Donaldson J, Liddle H, Titus JC, Kaminer Y, Webb C, Hamilton N, Funk R. The Cannabis Youth Treatment (CYT) Study: Main findings from two randomized trials. *Journal of Substance Abuse Treatment*. 2004; 27:197–213. [PubMed: 15501373]
- Dennis ML, Titus JC, Diamond G, Donaldson J, Godley SH, Tims F, Webb C, Kaminer Y, Babor T, Roebuck MC, Godley MD, Hamilton N, Liddle H, Scott CK. The Cannabis Youth Treatment (CYT) experiment: Rationale, study design, and analysis plans. *Addiction*. 2002; 97:S16–S34.
- Garner BR. Research on the diffusion of evidence-based treatments within substance abuse treatment: A systematic review. *Journal of Substance Abuse Treatment*. (in press).
- Garner BR, Godley SH, Funk RR, Dennis ML, Smith JE, Godley MD. Exposure to A-CRA treatment procedures as a mediator of the relationship between adolescent substance abuse treatment retention and outcome. *Journal of Substance Abuse Treatment*. 2009; 36:252–264. [PubMed: 18715742]
- Godley, SH.; Meyers, RJ.; Smith, JE.; Godley, MD.; Titus, JM.; Karvinen, T.; Dent, G.; Passetti, L.; Kelberg, P. DHHS Publication No. (SMA) 01–3489. Rockville, MD: Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration; 2001. Cannabis Youth Treatment Series Volume 4: The Adolescent Community Reinforcement Approach.
- Henggeler, SW.; Borduin, CM. Multisystemic therapy adherence scales. Charleston: Medical University of South Carolina, Department of Psychiatry and Behavioral Science; 1992.
- Henggeler SW, Melton GB, Brondino MJ, Scherer DG, Hanley JH. Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology*. 1997; 65:821–833. [PubMed: 9337501]
- Henggeler, SW.; Schoenwald, SK.; Borduin, CM.; Rowland, MD.; Cunningham, PB. Multisystemic treatment of antisocial behavior in children and adolescents. New York: Guilford Press; 1998.

- Higgins ST, Delaney DD, Budney AJ, Buckel WK, Hughes JR, Foerg F, Fenwick JW. A behavioral approach to achieving initial cocaine abstinence. *American Journal of Psychiatry*. 1991; 148:1218–1224. [PubMed: 1883001]
- Hogue A, Henderson CE, Dauber S, Barajas PC, Fried A, Liddle HA. Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology*. 2008; 76:544–555. [PubMed: 18665684]
- Hunt GM, Azrin NH. A community-reinforcement approach to alcoholism. *Behavior Research and Therapy*. 1973; 11:91–104.
- Institute of Medicine IOM. Bridging the gap between practice and research: Forging partnerships with community-based drug and alcohol treatment. Washington, DC: National Academy Press; 1998.
- Madson MB, Campbell TC, Barrett DE, Brondino MJ, Melchert TP. Development of the motivational interviewing supervision and training scale. *Psychology of Addictive Behaviors*. 2005; 19:303–310. [PubMed: 16187810]
- Miller WR, Sorensen JL, Selzer JA, Brigham GS. Disseminating evidence-based practices in substance abuse treatment: A review with suggestions. *Journal of Substance Abuse Treatment*. 2006; 31:25–39. [PubMed: 16814008]
- Moyers TB, Martin T, Manuel JK, Hendrickson SML, Miller WR. Assessing competence in the use of motivational interviewing. *Journal of Substance Abuse Treatment*. 2005; 28:19–26. [PubMed: 15723728]
- Orwin RG. Assessing program fidelity in substance abuse health services research. *Addiction*. 2000; 95:S309–S328. [PubMed: 11132360]
- Pierson HM, Hayes SC, Gifford EV, Roget N, Padilla M, Bissett R, Berry K, Kohlenberg B, Rhode R, Fisher G. An examination of the motivational interviewing treatment integrity code. *Journal of Substance Abuse Treatment*. 2007; 32:11–17. [PubMed: 17175394]
- Sechrest L, West SG, Phillips MA, Redner R, Yeaton W. Introduction. Some neglected problems in evaluation research: Strength and integrity of treatments. *Evaluation Review Studies Annual*. 1979; 4:15–38.
- Smith, JE.; Lundy, SL.; Gianini, L. Community reinforcement approach (CRA) and adolescent community reinforcement approach (A-CRA) coding manual. Albuquerque, NM: University of New Mexico; 2007.
- Smith JE, Meyers RJ, Delaney HD. The community reinforcement approach with homeless alcohol-dependent individuals. *Journal of Consulting and Clinical Psychology*. 1998; 66:541–548. [PubMed: 9642893]
- Waltz J, Addis ME, Koerner K, Jacobsen NS. Testing the integrity of psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*. 1993; 61:620–630. [PubMed: 8370857]
- Weisner, C.; McLellan, T.; Barthwell, A.; Blitz, C.; Catalano, R.; Chalk, M.; Chinnia, L.; Collins, RL.; Compton, W.; Dennis, ML.; Frank, R.; Hewitt, W.; Inciardi, JA.; Lightfoot, M.; Montoya, I.; Sterk, CE.; Wood, J.; Pintello, D.; Volkow, M.; Michaud, SE. Report of the Blue Ribbon Task Force on Health Services Research at the National Institute on Drug Abuse. Rockville, MD: National Institute on Drug Abuse; 2004.
- Yeaton WH, Sechrest L. Critical dimensions in the choice of maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology*. 1981; 49:156–167. [PubMed: 7217482]