Journal of Consulting and Clinical Psychology
2008, Vol. 76, No. 4, 531–543

0022-006X/08/$12.00 DOI: 10.1037/0022-006X.76.4.531

Outpatient Interventions for Adolescent Substance Abuse: A Quality of Evidence Review

Sara J. Becker and John F. Curry
Duke University

Previous reviews of outpatient interventions for adolescent substance abuse have been limited in the extent to which they considered the methodological quality of individual studies. The authors assessed 31 randomized trials of outpatient interventions for adolescent substance abuse on 14 attributes of trial quality. A quality of evidence score was calculated for each study and used to compare the evidence in support of different outpatient interventions. Across studies, frequently reported methodological attributes included presence of an active comparison condition, reporting of baseline data, use of treatment manuals, and verification of self-reported outcomes. Infrequently reported attributes included power and determination of sample size, techniques to randomize participants to condition, specification of hypotheses and primary outcomes, use of treatment adherence ratings, blind assessment, and inclusion of dropouts in the analysis. Treatment models with evidence of immediate superiority in 2 or more methodologically stronger studies included ecological family therapy, brief motivational interventions, and cognitive–behavioral therapy.

Keywords: adolescent substance abuse, methodological quality, outpatient interventions

Substance abuse among adolescents remains a widespread, persistent, and recurring public health problem. Relative to adult substance abusers, adolescent substance abusers experience a unique constellation of problems including more rapid progression from first use to abuse or dependence, shorter time from first to second dependence diagnosis, and more co-occurring psychiatric problems (Clark, Kirisci, & Tarter, 1998; Winters, 1999). Higher rates of substance use are also associated with leading causes of death among youths: accidents, suicide, and violent crime (Windle & Davies, 1999). Despite the serious and potentially lethal outcomes associated with adolescent substance abuse, less than 10% of adolescents in the community who meet criteria for substance abuse or dependence receive treatment (Substance Abuse and Mental Health Services Administration [SAMHSA], 2007). Moreover, data from national treatment admissions and discharges indicate that over 50% of those who receive treatment in the community drop out or terminate with unsatisfactory progress (SAMHSA, 2004).

The pervasive consequences and modest community outcomes associated with adolescent substance use require the identification of effective treatment models. Evaluating outpatient models is especially important since over 80% of adolescents receive treatment in outpatient or intensive outpatient settings (SAMHSA, 2004). Previous reviews of outpatient interventions have identified several effective treatment models but have been limited in the extent to which they evaluated the quality of evidence in support of these models. There is increasing evidence that inadequate methodological approaches and reporting are associated with overestimation of treatment effects (Juni, Altman, & Egger, 2001; Schulz, Chalmers, Hayes, & Altman, 1995). Failure to consider the quality of methodological design and reporting thereby limits the ability to detect potentially inflated treatment estimates, identify sources of bias, characterize strengths and weaknesses, and inform best practices in the field (see Moyer & Finney, 2005). Indeed, the assessment of methodological quality has been identified as one of the most important steps of the peer review process (Kassirer & Campion, 1994) and as one of the most critical components of systematic reviews (Chalmers et al., 1981).

The most comprehensive review of adolescent treatment outcome research to date was written by Williams, Chang, and the Addiction Centre Adolescent Research Group (2000). This review identified 53 studies, of which only 14 were controlled comparisons. When discussing these studies, the authors frequently referred to methodologically strong and weak designs, without a clear scheme as to how the quality of evidence was assessed. Similarly, the authors aptly highlighted the heterogeneity of prior research methods, but they did not provide guidance to synthesize the results. Thus, there is a need for a more systematic and current analysis of both the quality and level of evidentiary support for different outpatient interventions.

Other notable reviews on treatment outcomes for adolescent substance use include those by Waldron and Kaminer (2004); Deas and Thomas (2001); Muck and colleagues (2001); Ozechowski and Liddle (2000); Stanton and Shadish (1997); and Waldron

Sara J. Becker and John F. Curry, Department of Psychology and Neuroscience, Duke University.

A portion of this manuscript is based on a major area paper completed by Sara J. Becker under the guidance of John F. Curry, in partial fulfillment of requirements for a master of arts degree in the clinical psychology doctoral program at Duke University. The authors would like to thank Patricia Simon for her assistance in piloting an earlier version of the rating scheme used in this article.

Correspondence concerning this article should be addressed to Sara J. Becker, Box 90085, Department of Psychology and Neuroscience, Duke University, Durham, NC 27708. E-mail: sjb21@duke.edu

(1997). Several of the reviews (Ozechowski & Liddle, 2000; Stanton & Shadish, 1997; Waldron, 1997) focused exclusively on outcomes of family-focused models, while another (Waldron & Kaminer, 2004) focused on evidence in support of cognitive–behavioral therapy (CBT). Only two reviews discussed evidence in support of multiple treatment models. Deas and Thomas (2001) reviewed the outcomes of both pharmacotherapy and psychotherapy trials, while Muck and colleagues (2001) summarized the research in support of behavioral, family-oriented, and 12-step programs. The review of family and couples therapy by Stanton and Shadish (1997) was the only one to account for the methodological quality of different trials in the synthesis of results; in this meta-analysis, each study was assigned a quality score by using a 30-point design quality rating scale. None of these reviews considered the methodological quality of the evidence base in support of different treatment models.

A recent synthesis by Vaughn and Howard (2004) addressed important limitations of prior reviews by comparing treatment outcomes across different types of interventions. Similar to the approach taken by Stanton and Shadish (1997), this review assessed the methodological quality of the identified studies by using a composite rating scale. Moreover, the authors considered the methodological quality ratings when discussing the evidence in support of the different interventions. Results of this review suggested that multidimensional family therapy and cognitive–behavioral group treatment were the outpatient models with the highest level of evidentiary support.

Notwithstanding these contributions, the 16-point methodological rating scale used by Vaughn and Howard (2004) did not consider several issues germane to the study of adolescent substance use outcomes, and to the study of psychotherapy outcomes more broadly, such as the procedures used to randomize participants, the nature of the control condition, and the use of statistical techniques to handle attrition. Furthermore, the rating scheme relied on a composite score without providing a working definition of methodological quality or indicating the specific attributes met by individual studies. There is growing consensus that reliance on numeric rating scales alone is problematic due to significant variability across scales, limited data on reliability, susceptibility to reviewer bias, and inadequate differentiation of multiple components of trial quality (Brouwers et al., 2005; Juni, Witschi, Bloch, & Egger, 1999; Moja et al., 2005). Trial quality is a multidimensional construct that could pertain to a trial's internal validity, external validity, clinical relevance, or the quality of the reporting (Moyer & Finney, 2005). Internal validity refers to the degree to which differences between groups allocated to different interventions can be attributed to the interventions under investigation, whereas external validity reflects the extent to which the results of a trial are generalizable to other circumstances (Cook & Campbell, 1979). Internal validity is a necessary, but not sufficient, condition for external validity; if a study is methodologically flawed then the results lack validity and the question of generalizability becomes less significant. Consequently, a number of researchers have recommended using evaluation schemes that prioritize individual aspects of study design, focus on criteria that have been shown to influence internal validity, and establish replicable coding procedures (Juni et al., 1999; Moja et al., 2005; Moyer & Finney, 2005).

Consistent with these recommendations, the present review aims to apply individual attributes of trial quality to evaluate the evidence base in support of different outpatient interventions for adolescent substance abuse. Our primary objective is to evaluate randomized trials of outpatient interventions by using individual trial attributes that have been well established conceptually or empirically as influencing methodological quality. In line with the conceptualization endorsed by Moyer and Finney (2005), we define methodological quality as reflecting the extent to which the design, conduct, and analysis of trial data optimize internal validity by minimizing selection, measurement, and confounding bias on intervention effect estimates. Our secondary objective is to identify the level of evidentiary support for different outpatient interventions for adolescent substance abuse. Pursuit of these objectives represents an important step toward highlighting methodological attributes that have been shown to influence intervention effect size estimates, toward characterizing strengths and weaknesses in trials of adolescent substance use interventions, and toward identifying promising interventions that are supported by sound methodological evidence.

Method

Study Selection

Studies of outpatient interventions were selected according to the following criteria: (a) randomized design comparing an active outpatient treatment with one or more comparison conditions; (b) average age of participants between 13 and 18 years; (c) participants selected because of elevated levels of substance use or formal diagnosis of substance abuse or dependence; (d) intervention intended by the investigators to target substance use; (e) trial designed to evaluate treatment outcomes and not only elements of therapeutic process; and (f) published in a journal in English.

Search Strategy

The literature search sought to identify all randomized evaluations of outpatient treatments for adolescent substance abuse or dependence published in peer-reviewed journals through June 2007. Given our focus on the quality of trial design and reporting, we included only randomized trials that had already passed a process of peer review and did not include trials in book chapters, doctoral dissertations, or unpublished manuscripts. Studies were identified through: (a) database searches of MEDLINE (1966–2007) and PsycInfo (1966–2007); (b) examination of reference lists of relevant review articles and retrieved articles; (c) manual search of all issues from 1966 to 2007 of the one journal (*Journal of Consulting and Clinical Psychology*) in which more than 3 randomized evaluations were identified by our search. Database searches used combinations of substance terms (*substance, drug, alcohol, marijuana,* or *cannabis*), diagnosis or problem terms (*use, abuse,* or *dependence*), treatment terms (*treatment, intervention, therapy, psychotherapy,* or *counseling*), and design terms (*random, randomized, randomization, control, controlled,* or *trial*). Trials represented the unit of analysis such that separate articles published from the same dataset (e.g., cases in which posttreatment and follow-up results were reported separately) were treated as a single trial, whereas multiple trials reported in the same publication were treated as separate trials. The search identified a total of 31 trials across 30 publications.

Selection of Methodological Criteria

Attributes with a well-established link to internal validity were selected by using a three-step process that considered accepted guidelines for randomized controlled trials, precedence in adolescent substance abuse studies, and recent methodological and statistical advances in the field. As a first step, we reviewed the methodological guidelines for the design and reporting of randomized controlled trials recommended in the CONSORT (*Con*solidated *S*tandards of *R*eporting *T*rials) statement (Moher, Schultz, & Altman, 2001). The CONSORT statement was created to facilitate critical appraisal and evaluation of randomized trials by providing common guidelines for authors to improve the quality of trial reporting. Since its introduction in 1996, CONSORT has been adopted by an increasing number of medical, psychiatric, and clinical psychology journals (Altman, 1996; Dulcan, 2001; Huston & Hoey, 1996; La Greca, 2005) as well as by the Publications Board of the American Psychological Association, thereby improving the quality of trial reporting across disciplines (Plint et al., 2006).

The current version of the CONSORT (Moher et al., 2001) consists of a checklist of 22 evidence-based items that should be included in manuscripts of randomized controlled trials. The following 10 CONSORT items pertain specifically to the trial methods: *participants, interventions, objectives, outcomes, sample size, sequence generation, allocation concealment, implementation of randomization, blinding,* and *adequate statistical methods.* Three of these attributes were excluded due to redundancy with our selection of studies that specified inclusion criteria (*participants*), focus on quality of reporting as opposed to internal validity (*implementation of randomization*), and focus on analytical methods beyond the scope of our review objectives (*statistical methods*). In addition, to increase relevance to psychotherapy research, the attribute *blinding* (which broadly pertains to whether the participants, those administering the interventions, and those evaluating the outcomes are blind to group assignments) was modified to focus specifically on *blind assessment* by an independent evaluator.

Next, we augmented the seven methodological attributes retained from the CONSORT statement with attributes that have been used to evaluate the quality of adolescent substance abuse studies. A focused search identified one quality assessment scheme, the Methodological Quality Rating Scale (MQRS; Miller et al., 1995), that was specifically designed to evaluate trials of substance abuse interventions. The MQRS consists of 13 dimensions of trial quality pertaining to the study design, analytical strategy, and collection of measures: *study design, replicability of procedures, baseline data, quality control of intervention, follow-up length, dosage, collateral report, objective verification, dropouts/attrition enumerated, statistical power, independent evaluator, statistical analyses,* and *multisite replications*.

Five of the MQRS attributes were excluded due to redundancy with our focus on randomized trials (*study design*), redundancy with attributes included in the CONSORT statement (*statistical analyses, independent evaluator*), and focus on durability or external validity of results as opposed to internal validity (*follow up length, multisite replication*). Another four attributes were excluded on the basis of how frequently they were reported in a recent review of controlled studies by Vaughn and Howard (2004). As noted by the developers of the Jadad scale, a well-validated quality scale in the medical literature (Jadad et al., 1996), a methodological attribute does not provide much power to differentiate among studies if it is endorsed by more than 85% or less than 15% of studies. By using a slightly more conservative criterion, we removed four attributes that were reported in over 90% of studies included in Vaughn and Howard's (2004) review: *replicability of procedures, quality control of intervention, dosage,* and *dropouts/attrition enumerated.*

As a final step, we modified the list of 11 attributes to capture several critical design variables that reflect recent advances in clinical intervention research. One key methodological feature that has been empirically shown to affect treatment outcomes is whether the intervention of interest is compared with an active or passive comparison condition. Active conditions control for expectancy effects as well as for improvement due to nonspecific factors (e.g., attention or therapeutic relationship), whereas passive conditions control only for the passage of time and the natural progression of symptoms and disorders. It is well documented that comparisons with a passive condition are associated with larger treatment effects than are comparisons of two active treatments (Baskin, Tierney, Minami, & Wampold, 2003; Kazdin, Bass, Ayers, & Rodgers, 1990; Weisz, McCarty, & Valeri, 2006); thus, we included the presence of an *active comparison* as an attribute indicative of a higher quality of evidence in support of an intervention.

In recent years, two additional design features that have received recognition include treatment manuals and treatment adherence ratings. Both techniques have been advocated as means to standardize procedures and promote adherence, thereby increasing the likelihood that observed treatment outcomes are attributable to the specific intervention (Lonigan, Elbert, & Johnson, 1998). Higher levels of treatment adherence have been associated with more favorable and less variable treatment outcomes (Luborsky, McLellan, Diguer, Woody, & Seligman, 1997; Waltz, Addis, Koerner, & Jacobson, 1993), highlighting the importance of these design features in protecting internal validity. The CONSORT statement encompasses these attributes under the general *intervention* attribute, which pertains to whether details about the intervention and its delivery are provided for each treatment group. Given the accepted value of treatment manuals and tests of therapist adherence in promoting consistent treatment delivery (Waltz et al., 1993), we replaced the CONSORT's general *intervention* attribute with two specific attributes: *manualized treatment* and *treatment adherence ratings*.

A final characteristic that has been widely recognized as protecting against biased treatment estimates is the inclusion of dropouts in the statistical analysis (Howard, Krause, & Orlinksky, 1986; Stanton & Shadish, 1997). Although the CONSORT and MQRS consider whether dropouts are enumerated and discussed, neither differentiates between trials that report attrition levels and those that account for attrition statistically in the analysis. In recent years, the application of intention-to-treat (ITT) analysis has allowed attrition to be accounted for in the analysis by using all available data from participants randomized into a trial. A full ITT analysis (Hill, 1961) requires that data from all participants randomized to treatment be utilized, regardless of two different types of attrition: (a) attrition from the treatment and (b) attrition from the research assessments. Analyses that include only participants

who completed a portion of the treatment are defined as *treated case* analyses, while analyses that include only participants who completed the research assessments are defined as *available case* analyses. While available case analyses typically exclude fewer participants and therefore provide greater protection against bias than do treated cases analyses, ITT analysis is the only approach that unequivocally produces consistent and unbiased treatment outcomes (Ellenberg, 1996). We therefore included *ITT analysis* as a three-level attribute, with differentiation among studies using treated case, available case, and ITT analysis.

Ultimately, we applied 14 attributes to evaluate the methodological strength of the 31 randomized trials: 6 from the CONSORT statement, 4 from the MQRS, and 4 reflecting recent advances in the field. Table 1 displays the rating criteria for these 14 dimensions of methodological quality.

Rating Procedure

In harmony with recent recommendations (Brouwers et al., 2005; Juni et al., 2001; Moja et al., 2005), published randomized controlled trials were rated with regard to individual attributes of trial quality with an established link to internal validity. The focus on published trials presumably introduced bias in favor of methodologically stronger studies. Study co-authors independently reviewed 31 trials for the presence of each attribute, with reliability assessed by using Cohen's weighted kappa coefficient. Discrepancies were resolved through the coders' joint review.

To facilitate relative comparisons across studies, we augmented consideration of individual attributes with the use of a composite quality of evidence score (QES) to indicate the number of methodological attributes met. Each of the 14 methodological attributes received a rating of 0 (*not met or unclear*) or 1 (*met*), with the exception of ITT analysis, which received a possible rating of 0 (*treated case analysis*), 1 (*available case analysis*), or 2 (*full ITT analysis*). Ratings of the 14 individual attributes were added to calculate total QES, with a possible range from 0 to 15. Studies were then dichotomously classified, with studies having a QES below the median referred to as methodologically weaker and studies having a QES equal to or above the median referred to as methodologically stronger. Results of the studies were also dichotomously coded as having positive or neutral/negative results, with

Table 1
Individual Attributes of Methodological Quality

Attribute		Criteria for rating
1. Objective	1=	Specific objectives and hypotheses. Hypotheses are amenable to explicit statistical evaluation.
	0=	Objectives or hypotheses not explicitly established.
2. Sample size	1=	Process for determining the sample size discussed along with any interim analyses and stopping rules.
	0=	Determination of sample size not discussed.
3. Power	1=	Study is sufficiently powered to detect differences between treatment groups (e.g., at least 71 subjects per condition with active comparison, 27 subjects per condition with passive comparison).
	0=	Study is not sufficiently powered.
4. Outcome	1=	Established primary and secondary outcome measures. Primary outcome is specified as outcome of greatest importance.
	0=	Primary or secondary outcome measures are not specified.
5. Sequence generation	1=	Process for generating a random sequence described with sufficient detail to confirm that each participant had an unpredictable, independent chance of receiving each intervention.
	0=	Process was not purely random, unspecified.
6. Allocation concealment	1=	Process of assigning participants to groups described with sufficient detail to confirm that investigators recruiting and conducting the initial assessment could not discern the participant's treatment group.
	0=	Process was not concealed, unspecified.
7. Active comparison	1=	At least one active comparison (e.g., alternate model, treatment as usual).
	0=	All comparison conditions were passive (e.g., waitlist, no-treatment control).
8. Baseline data	1=	Baseline demographic and clinical characteristics reported by condition.
	0=	Baseline demographic or clinical characteristics not reported.
9. Manualized treatment	1=	At least one treatment condition was guided by a manual.
	0=	None of the treatments were guided by a manual, unspecified.
10. Treatment adherence rating	1=	Treatment adherence monitored with scales, checklists, or rating forms completed by therapist, supervisor, independent observer, and/or patient.
	0=	Treatment adherence was not monitored using rating forms, unspecified.
11. Collateral report	1=	At least one outcome is a collateral report (e.g., parent, caregiver, teacher).
	0=	No collateral report.
12. Objective measure	1=	At least one outcome is an objective measure (e.g., urine, blood samples, paper records).
	0=	No objective verification.
13. Intention-to-treat (ITT) analysis	2=	ITT analysis. All subjects analyzed in groups to which they were assigned.
	1=	Available case analysis. Only subjects who completed one or more research assessments were analyzed.
	0=	Treated case analysis. Only subjects who completed a portion of the treatment were analyzed.
14. Blind assessment	1=	Follow-up assessments completed by treatment-blind evaluator.
	0=	Follow-up not completed by blind evaluator, unspecified.

indication of superior immediate outcomes relative to a comparison condition considered as positive findings in support of an intervention.

Results and Discussion

Methodological Attributes Across Trials

The kappa value and proportion of studies meeting each of the 14 methodological criteria are depicted in Table 2. We initially reached a modest level of agreement on one attribute, *collateral report* ($\kappa = 0.53$), because of differences in our interpretation of whether the collateral report pertained narrowly to the adolescent's self-reported substance use or broadly to any adolescent-reported dependent variable. We ultimately decided to use the more broad definition, because the resulting rate of studies meeting this attribute was more consistent with the rate found in Vaughn and Howard's (2004) review. A third independent rater then reviewed the 31 studies by using the clarified definition, resulting in a kappa value of 1.00 with the original rater who had used this definition. Kappa values of the remaining 13 attributes ranged from 0.80 to 1.00, indicating substantial to perfect agreement (Landis & Koch, 1977).

Methodological attributes met by each of the 31 studies are presented in Table 3. Areas of relative methodological strength, present in over 50% of studies, included the following five attributes: *active comparison, baseline data, manualized treatment, collateral report,* and *objective measures*. In contrast, areas of relative weakness, reported in less than 50% of studies, included the following nine attributes: *objectives, sample size, power, outcomes, sequence generation, allocation concealment, treatment adherence ratings, ITT analysis,* and *blind assessment.*

Most Frequently Reported Attributes

Active comparison represented the most frequently reported attribute, occurring in over 80% of the studies ($n = 25$). Most studies using an active condition directly compared one or more prespecified interventions ($n = 21$), although several ($n = 4$) compared an active treatment with treatment as usual. Active comparison conditions varied significantly in dosage, with interventions ranging from merely the provision of referrals to the provision of over 12 weeks of treatment. Passive comparison conditions included no treatment ($n = 4$) and delayed treatment ($n = 2$). The small proportion of studies using a passive or no-treatment comparison group was consistent with prior reviews (Vaughn & Howard, 2004; Williams et al., 2000) and likely reflected ethical concerns about the use of a no-treatment comparison as well as a general consensus in the adolescent substance abuse field that some treatment is better than no treatment (Catalano, Hawkins, Wells, Miller, & Brewer, 1991). While the use of an active control provides a more stringent test of a treatment's effectiveness, the lack of a passive control makes it difficult to interpret the results of multiple studies (32%; $n = 10$) in which the active treatments resulted in an equivalent level of improvement across all time points. Interpretations were rendered even more complex when studies did not provide background evidence confirming that the active comparison intervention had previously demonstrated effectiveness relative to no or minimal treatment. Challenges in interpreting the results of active comparisons were noted by the investigators of several studies ($n = 7$) included in the present analysis.

Another frequently met attribute was *baseline data*, which requires reporting of demographic and clinical characteristics by treatment group. Reporting of baseline clinical characteristics occurred more often than did reporting of key demographic characteristics (e.g., gender, age, race). Approximately 94% of the studies ($n = 29$) reported raw data on at least one clinical characteristic by group, a rate that was somewhat higher than the 87% of studies reporting baseline clinical data in Vaughn and Howard's (2004) review. Fewer studies (77%; $n = 24$) reported that the groups were equivalent on demographic factors. All of the studies that reported equivalence of demographic factors also reported raw data on the clinical characteristics, resulting in an overall rate of 77% studies ($n = 24$) meeting this attribute.

Other areas of methodological quality that were frequently reported included techniques used to verify self-reported outcomes. Verification via *collateral report* and *objective measures* occurred in 61% ($n = 19$) and 65% ($n = 20$) of studies, respectively. Relative to the results of Vaughn and Howard's (2004) review, the rate of collateral interviews was highly consistent (61% vs. 60%), whereas the rate of objective verification was somewhat lower (65% vs. 73%). Overall, 81% ($n = 25$) of the studies used either collateral interviews or objective measures, and 45% ($n = 14$) used both techniques.

A final area of methodological strength was the use of *manualized treatment*, which was met by 81% of studies ($n = 25$). By contrast, the other technique to promote treatment fidelity, *treatment adherence ratings*, was reported by only 48% of studies ($n = 15$). Eleven of the 25 studies that used a manual did not obtain adherence ratings, whereas only 2 of the 15 studies that obtained treatment ratings did not explicitly reference the use of a manual. Of the 11 manualized studies that did not obtain adherence ratings, the majority ($n = 8$) specifically mentioned the provision of ongoing supervision to promote fidelity. Combined, 87% of the 31 studies ($n = 27$) used at least one of these techniques to promote adherence, and 42% ($n = 13$) used both approaches.

Table 2
Methodological Attributes Across Randomized Trials of Adolescent Substance Abuse Interventions (N = 31)

Methodological attribute	*k*	*N*	%
1. Objective	0.87	13	42
2. Sample size	1.00	2	6
3. Power	0.93	7	23
4. Outcome	0.80	6	19
5. Sequence generation	0.92	8	26
6. Allocation concealment	0.80	4	13
7. Active comparison	0.90	25	81
8. Baseline data	0.85	24	77
9. Manualized treatment	0.83	25	81
10. Treatment adherence rating	0.94	15	48
11. Collateral report	1.00	19	61
12. Objective measure	0.87	20	65
13. Intention-to-treat analysis	0.82	10	32
14. Blind assessment	0.80	5	16

Table 3
Attributes of Randomized Trials of Outpatient Interventions for Adolescent Substance Abuse (N = 31 Trials)

Author, year (QES)	Objective	Baseline data	Sample	Power	Outcome	Active	Random sequence	Allocation concealed	Manual	Tx ratings	Collateral report	Objective measure	ITT	Blind assess
1. Azrin et al., 1994 (5)	No	Yes	No	No	No	Yes	No	No	No	Yes	Yes	Yes	TC	No
2. Azrin et al., 2001 (7)	No	Yes	No	No	No	Yes	No	No	Yes	Yes	Yes	Yes	TC	Yes
3. Bailey et al., 2004 (4)	Yes	No	No	No	No	No	No	No	Yes	No	No	No	ITT	No
4. Conrad et al., 2006 (6)	Yes	Yes	No	Yes	No	No	Un	Un	Yes	No	No	No	ITT	No
5. Dennis et al., 2004, Experiment 1 (12)	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	AC	Yes
6. Dennis et al., 2004, Experiment 2 (12)	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	AC	Yes
7. Friedman, 1989 (6)	No	Yes	No	No	No	Yes	Un	Un	Yes	Yes	Yes	No	AC	Un
8. Henggeler et al., 1999 (7)	No	Yes	No	No	No	Yes	Un	Un	Yes	Yes	No	Yes	ITT	No
9. Henggeler et al., 2006 (8)	No	Yes	No	No	No	Yes	Un	Un	Yes	Yes	Yes	Yes	ITT	Un
10. Joanning et al., 1992 (4)	No	No	No	No	No	Yes	No	No	Yes	No	Yes	Yes	TC	No
11. Kaminer et al., 1998 (8)	Yes	Yes	No	No	No	Yes	Un	Un	Yes	Yes	Yes	Yes	AC	Un
12. Kaminer et al., 2002 (7)	Yes	Yes	No	No	No	Yes	Un	Un	Yes	No	No	Yes	ITT	Un
13. Latimer et al., 2003 (6)	No	Yes	No	No	No	Yes	Un	Un	Yes	No	Yes	Yes	AC	Un
14. Lewis et al., 1990 (4)	No	No	No	No	Yes	Yes	Un	Un	No	No	Yes	Yes	TC	No
15. Liddle et al., 2001 (6)	Yes	Yes	No	No	No	Yes	Un	Un	Yes	No	Yes	Yes	TC	Un
16. Liddle et al., 2004 (9)	Yes	Yes	No	No	No	Yes	Yes	Un	Yes	Yes	Yes	No	Un	Yes
17. Marsden et al., 2006 (11)	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes	ITT	No
18. McCambridge & Strang, 2004 (5)	No	Yes	No	Yes	No	No	Un	Un	Yes	No	No	No	ITT	No
19. McGillicuddy et al., 2001 (4)	No	Yes	No	No	No	No	No	No	No	Yes	No	No	ITT	Un
20. Najavatis et al., 2006 (8)	No	Yes	Yes	No	Yes	Yes	Un	Un	Yes	Yes	No	No	ITT	No
21. Peterson et al., 2006 (7)	Yes	Yes	No	Yes	No	No	Yes	Un	Yes	No	No	Yes	AC	Un
22. Santisteban et al., 2003 (7)	Yes	Yes	No	No	No	Yes	Un	Un	Yes	Yes	Yes	Yes	TC	Un
23. Slesnick & Prestopnik, 2005 (7)	Yes	Yes	No	No	No	Yes	Yes	Un	Yes	No	No	Yes	AC	Un
24. Smith et al., 2006 (6)	Yes	No	No	No	No	Yes	Un	Un	Yes	No	Yes	Yes	AC	Un
25. Srisurapanont et al., 2007 (7)	No	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes	TC	No
26. Szapocznik et al., 1983 (2)	No	No	No	No	No	Yes	Un	Un	No	No	Yes	No	TC	Un
27. Szapocznik et al., 1986 (2)	No	No	No	No	No	Yes	Un	Un	No	No	Yes	No	TC	Un
28. Thush et al., 2006 (5)	No	No	No	No	No	Yes	Un	Un	No	No	Yes	Yes	ITT	No
29. Waldron et al., 2001 (10)	Yes	Yes	No	No	Yes	Yes	Yes	Un	Yes	Yes	Yes	Yes	AC	Un
30. Walker et al., 2006 (5)	No	Yes	No	Yes	No	No	Un	Un	Yes	Yes	No	No	AC	Un
31. Winters & Leitten, 2007 (7)	Yes	Yes	No	No	No	Yes	Un	Un	Yes	Yes	Yes	No	TC	Yes

Note. QES = quality of evidence score; Tx = treatment; ITT = intention-to-treat analysis; TC = treated case analysis; AC = available case analysis; Un = unclear.

Least Frequently Reported Attributes

Two of the least frequently reported attributes pertained to the determination of an adequate sample size. Sample size considerations include whether the study provided an a priori justification of the *sample size* and whether the study had sufficient *power* to test for differences between treatment groups. According to the general guidelines recommended by Kazdin and Bass (1989), treatment studies with an active comparison condition require at least 71 participants per group to have adequate power, while studies with a passive comparison condition require at least 27 participants per group. With these guidelines, 77% of studies (n = 24) lacked sufficient power to test for differences between conditions. Inadequate power was compounded by a lack of reporting about the determination of the sample size; approximately 95% of studies (n = 29) neglected to state how the sample size was determined. Within the current sample, approximately 1 in 4 studies (23%; n = 7) concluded that there was no significant difference between two or more treatment conditions, despite having insufficient power to support such a claim.

Two other infrequently reported areas pertained to the process of randomization, which has been identified as the critical component of high-quality randomized trials (Schulz & Grimes, 2002). The CONSORT statement defines adequate randomization as encompassing two independent, yet related, processes: generation of an unpredictable allocation sequence (*sequence generation*) and concealment of this sequence from the investigators responsible for enrolling participants (*allocation concealment*). The first process, *sequence generation*, refers to the use of a random, probabilistic sequence to assign participants. Methods such as alternate assignment, case record ID, date of birth, or date of enrollment are not random but, rather, are systemic occurrences. In accordance with guidelines offered by Schulz and Grimes (2002), we viewed sequence generation as adequate if the following auditable, a priori methods were described: computerized randomization, random number table, coin toss, shuffled deck. Following these guidelines, 19% of trials (n = 6) reported using randomization even though seemingly nonrandom sequence generation techniques were used, and 55% (n = 17) did not specify the method used for sequence generation. Virtually all of the studies with adequate sequence generation (26%; n = 8) used a computer to generate a random sequence.

The second process related to randomization, *allocation concealment*, refers to the concealment of the allocation schedule from those investigators responsible for enrolling participants. If a random sequence is generated, but the sequence is not adequately concealed, then decisions about enrollment may be influenced by foreknowledge of upcoming treatment assignments, potentially introducing selection bias. Empirical analyses have shown that studies with inadequate or unclear concealment yield biased estimates of treatment effects, with odds ratios exaggerated by 30% to 40% (Kjaergard, Villumsen, & Gluud, 2001; Schulz et al., 1995). In the current analysis, we awarded credit for allocation concealment only if the study also had random sequence generation, reflecting our view that investigators would be more likely to ascertain a nonrandom sequence. This criterion was somewhat more conservative than the criterion used in prior methodological reviews (Schulz, Chalmers, Grimes, & Altman, 1994). Half of the current studies with sufficient sequence generation (n = 4) provided adequate detail to confirm concealment. Techniques used to conceal allocation included the use of an independent research staff to generate the allocation sequence (n = 3) as well as the use of opaque, sealed envelopes (n = 1). An additional 13% of studies (n = 4) stated that allocation was concealed, but they either did not confirm generation of a random sequence or did not specify the steps taken to conceal upcoming assignments.

The reporting of study *objectives* and primary *outcomes* represented other areas for improvement. Failure to specify a hypothesis or any primary outcomes can introduce bias to the extent that multiple analyses of the same dataset increase the risk of false positive results (Tukey, 1977). Although 100% of studies reported study objectives, only 42% of studies (n = 13) reported an explicit, testable hypothesis. An even smaller proportion of studies (19%, n = 6) explicitly defined one or two primary outcomes. The remaining studies had three or more outcomes, none of which was delineated as primary.

Another infrequently reported area was *ITT analysis*. Approximately one third of studies met this criterion (32%, n = 10), with the remaining studies fairly evenly split between use of treated case analysis that excluded treatment dropouts (32%, n = 10) and use of available case analysis that excluded research dropouts (35%, n = 11). Consistent with the findings of Hollis and Campbell (1999), a notable proportion of studies (16%; n = 5) inaccurately reported use of ITT analyses, when they had actually conducted an available case analysis. Those studies that provided sufficient detail to confirm ITT analysis utilized the following statistical techniques: last value carried forward (n = 4), random regression (n = 1), latent growth modeling with maximum likelihood estimation (n = 1), 100% data retention on archival records (n = 2), and 100% participation at follow-up (n = 2). A subset of studies (19%; n = 6) closely approached ITT analysis by using data from 95% or more of the sample, while another study described the use of latent growth modeling, which often implies the use of maximum likelihood imputation, but did not provide sufficient detail to make a definitive judgment. Those studies that conducted ITT analysis on one or more outcomes were all published since 1999, highlighting the recency of this methodological attribute.

The final area for improvement was *blind assessment* of outcome by an independent evaluator. Across the 31 studies, 16% (n = 5) reported meeting this criterion, 35% (n = 11) reported not meeting this criterion, and the remaining 48% (n = 15) did not adequately report this criterion. Those studies that clearly did not fulfill this attribute either explicitly stated that the follow-up was conducted by the therapist or a nonblind evaluator (n = 4), relied on an outcome that incorporated nonblind therapist judgments (n = 2), or relied on self-report measures only (n = 5). By contrast, those studies that did not provide sufficient detail did not explicitly report whether the evaluator administering the assessment was blind to treatment condition. Of note, many of the studies that provided insufficient detail reported the following attributes: evaluator was not a member of the clinical staff (n = 10); independent rater scored a portion of the assessment (n = 5); and evaluator was blind to the study hypothesis (n = 1). These strategies help to limit interpretation bias at the time of the assessment. However, Juni and colleagues (1999) found that open assessment of outcome was associated with exaggerated effect sizes of about 35% in an analysis of meta-analytic reviews, highlighting

the need to attend to blind assessment of outcomes in clinical research.

Quality of Evidence in Support of Different Interventions

Across the 31 studies, 57 active outpatient interventions were tested. The median QES across trials was 7.0, with a range from 2.0 to 12.0. There was a modest positive association between methodological quality and year of publication ($r = .55$, $p < .01$), reflecting a tendency for more recent studies to use methodologically stronger designs. Treatment approaches fell into five main modalities: family-based therapy, brief motivational therapy, CBT, adolescent group therapy, and parent training. In addition, there were integrated models of CBT and motivation enhancement therapy as well as CBT and functional family therapy. Level of evidentiary support by treatment modality and specific models associated with each modality are summarized in Table 4.

Family-based therapy was the most frequently tested approach, with 22 models evaluated across 17 studies. Building on prior classifications (Ozechowski & Liddle, 2000; Waldron, 1997), family-based interventions were broadly sorted into five types of models: systems, behavioral, functional, ecological, and educational. Classifications were based on how the treatment addressed intrafamilial and extrafamilial influences in the development, maintenance, and modification of adolescent substance use. Systems approaches attempted to restructure problematic family interaction patterns associated with the adolescent's substance use, while behavioral approaches applied principles of operant and social learning within the family context to promote prosocial behaviors and reduce substance use. Models that integrated principles of both systems and behavioral approaches were viewed as functional approaches. Ecological models expanded the boundaries of treatment beyond the family and utilized individualized strategies to target adolescent substance use in the context of multiple interrelated, nested systems. Finally, family education models focused on providing psychoeducation to the family of the adolescent substance user.

Similar to the overall pool of studies, QES of studies evaluating family therapy ranged from 2.0 to 12.0, with a median of 6.0. Nine of the 17 studies used methodologically stronger designs. Of the family therapy approaches, ecological family therapy was evaluated the most often ($n = 7$) and was the only therapy evaluated in more than 1 methodologically stronger study ($n = 6$). Median QES of ecological approaches was 8.0 and ranged from 6.0 to 12.0. Specific ecological models tested in methodologically stronger studies included multisystemic therapy ($n = 2$), multidimensional family therapy ($n = 2$), family systems network ($n = 1$), and ecologically based family therapy ($n = 1$). Three of these studies found that ecological models, delivered alone or integrated with juvenile court services, had superior outcomes to other active treatment conditions (e.g., service as usual through a shelter, juvenile court services, adolescent group therapy). Meanwhile, 3 studies found that ecological models had comparable outcomes with those of treatment as usual in the community as well as those of integrated models of CBT and motivation enhancement therapy.

Brief motivational interventions represented the second most frequently investigated outpatient intervention. Treatment models were classified as brief motivational interventions if they consisted of one or two sessions targeted toward increasing the adolescent's motivation to reduce substance use. Seven brief motivational interventions were tested across six studies, four of which used methodologically stronger designs. Studies evaluating motivational interventions had a median QES of 7.0 and a range from 5.0 to 11.0. Three of the four methodologically stronger studies found favorable evidence for brief motivational interventions, whereas one study found no differences between a one-session intervention and the provision of information.

Although not the most often tested intervention ($n = 4$), CBT represented the intervention supported by the greatest proportion

Table 4
Quality of Evidence Scores by Treatment Modality

Treatment approach	Study numbers	Median QES	# QES ≥7	# QES ≥7, + results	Min, Max
Family-based therapy					
Systems model (FST, BSFT)	10, 22, 26, 27	3.0	1	1	2, 7
Behavioral model (FBT, SOFT)	1, 2, 24	6.0	1	0	5, 7
Functional model (FFT, PBFT)	7, 14, 29	6.0	1	1	4, 10
Ecological model (MDFT, MST, EBFT, FSN)	5, 6, 8, 9, 15, 16, 23	8.0	6	3	6, 12
Educational model (multifamily groups)	10, 15	5.0	0	0	4, 6
Brief motivational intervention	17, 18, 21, 25, 30, 31	7.0	4	3	5, 11
Cognitive behavioral therapy	2, 11, 12, 20	7.5	4	2	7, 8
Adolescent group therapy					
Interactional	10, 11, 15, 16, 22, 29	7.5	4	0	4, 10
Psychoeducational	12, 13, 25	7.0	2	1	6, 7
Parent skills training	7, 19	5.0	0	0	4, 6
Integrated model					
Cognitive behavioral + motivation enhancement	3, 4, 5, 6, 24, 28, 29	6.0	3	0	4, 12
Cognitive behavioral + functional family	13, 29	8.0	1	1	6, 10

Note. Study numbers refer to numbers from Table 3. QES = quality of evidence score; # QES ≥ 7 = number of studies with quality of evidence score equal to or above the median; + results = evidence of immediate superiority to at least one comparison condition; FST = family systems therapy; BSFT = brief strategic family therapy; FBT = family behavioral therapy; SOFT = strengths oriented family therapy; FFT = family functional therapy; PBFT = Purdue brief family therapy; MDFT = multidimensional family therapy; MST = multisystemic therapy; EBFT = ecological-based family therapy; FSN = family support network.

of methodologically stronger studies (100%; $n = 4$). CBT models explicitly aimed to modify cognitive processes, beliefs, individual behaviors or environmental reinforcers associated with the adolescent's substance use. Median QES of CBT studies was 7.5, with a narrow range from 7.0 to 8.0. Models of CBT included both individual ($n = 2$) and group modalities ($n = 2$). Two studies were supportive of CBT as having superior effectiveness to group interactional therapy and treatment as usual, while the other two studies found that CBT had comparable outcomes with group psychoeducation and family behavior therapy.

Eight models across seven studies explicitly integrated principles of CBT and motivation enhancement therapy. QES of the seven studies ranged from 4.0 to 12.0, with a median of 6.0. Methodologically stronger designs were used in three of the studies. In two of the methodologically stronger studies, integrated CBT and motivation enhancement had comparable effectiveness with ecological family and community reinforcement approaches. In the final study, an integrated CBT and motivation model was compared with three conditions: adolescent group therapy, functional family therapy, and integrated CBT and functional family therapy. At the posttreatment assessment, the integrated CBT model was inferior to the two family models and equivalent to adolescent group therapy. However, no significant differences were found among the four treatments at the 3-month follow-up.

Another integrated model tested in two studies was the combination of CBT and functional family therapy, which had a median QES of 8.0. One of the two studies used a methodologically stronger design (QES = 10.0). As previously discussed, the methodologically stronger study compared an integrated model of CBT and functional family therapy with three other conditions. Although the integrated family model was superior to two of the conditions at the posttreatment assessment, this superiority was not maintained at the 3-month follow-up.

Adolescent group therapy represented an active treatment model that typically served as the comparison condition, as opposed to the primary intervention of interest. Group therapy models were tested in nine studies, three of which were psychoeducational in nature and six of which were more interactional . The psychoeducational groups emphasized the provision of information about alcohol and other drugs, while the interactional groups emphasized exploration of perceived benefits and consequences of substance use, identification of interpersonal processes within the group, and practice with skills needed to reduce substance use. Six of the nine studies used methodologically stronger designs, reflecting a representative proportion of studies testing psychoeducational and interactional approaches. Although all six of the methodologically stronger studies used treatment manuals for at least one condition, only three of the studies used manuals for the group therapy condition. Results of five of the six methodologically stronger studies were not favorable toward adolescent group therapy. In these studies, adolescent group therapy was outperformed by group CBT, family systems therapy, ecological family therapy, and a brief motivational intervention. The remaining methodologically stronger study found that a psychoeducation group had comparable acute outcomes with those of group CBT.

The final treatment approach was parent skills training, which aimed to teach parents the skills needed to promote effective coping, problem solving, communication, and/or parental monitoring. Median QES was 5.0, with a range from 4.0 to 6.0. Parent skills training was the only approach tested in more than one study ($n = 2$) that was not evaluated in any studies using a methodologically stronger design.

Conclusion

The primary objective of the current review is to evaluate published, peer-reviewed randomized trials on individual attributes that have been empirically or conceptually established as preventing biased treatment outcomes. Our review of 31 trials on 14 individual attributes reveals that 9 of the attributes were reported by less than 50% of randomized trials, indicating several areas for improvement. Although a randomized design was a prerequisite for study selection, few of the studies adequately reported the techniques utilized to generate a random sequence or to conceal allocation schedules. Sample sizes were generally small and were rarely justified, thereby increasing the risk of Type II error (failing to find a significant treatment effect when a treatment effect may exist). Moreover, studies rarely established a priori hypotheses or primary outcomes, introducing risks associated with multiple outcomes analyses. Less than one in five studies reported adequate blinding of outcome assessment, and less than one in three included all participants in the analysis. These methodological issues have been similarly documented in other disciplines (Adetugbo & Williams, 2000; DerSimonian, Charette, McPeek, & Mosteller, 1982; Liberati, Himel, & Chalmers, 1986; Schulz et al., 1994; Thornley & Adams, 1998), highlighting the challenges associated with clinical trial research and the need for uniform reporting guidelines across fields. Also congruent with methodological surveys in other disciplines or journals (Hollis & Campbell, 1999), evaluation of methodological quality revealed some concerns with inaccurate reporting. For instance, 19% of studies reported random assignment even though they did not appear to use purely random methods, and 16% of studies reported the use of ITT analysis even though they excluded participants who dropped out of research assessments.

A secondary objective of the current review was to evaluate the quality of evidence in support of different treatment models. Models that had evidence of immediate treatment superiority in two or more methodologically stronger studies included ecological family therapy, brief motivational intervention, and CBT. While there has been a trend for the quality of evidence to increase over the past 2 decades in this challenging area of clinical research, our findings indicate that the improvement has not been equivalent across approaches. For instance, family therapy models were the most frequently tested, yet ecological family therapy was the only family approach tested in two or more studies using methodologically stronger designs. Furthermore, higher levels of methodological quality were not necessarily associated with stronger evidence in support of an intervention. As an example, adolescent group therapy models, other than group forms of CBT, were tested in six methodologically stronger studies but did not demonstrate any evidence of superiority relative to comparison conditions.

Limitations

Results of this review should be considered in the context of several limitations. A primary limitation inherent in most methodological surveys is that inadequate reporting can be confounded

with poorer quality study methodology. The methodological limitations highlighted in our review may reflect inadequate reporting, flawed methodology, or both. Because our aim was to appraise strengths and weaknesses in our field both in methodology and in reporting, we intentionally limited our scope to methodological factors as reported in peer-reviewed journal articles. We opted not to query authors about whether they had used, but not reported, higher quality methods. In a prior methodological review, Liberati and colleagues (1986) contacted principal investigators and found that inadequate reporting often reflected inadequate methods. Similarly, Schulz and colleagues (1995) found that inadequately conducted and inadequately reported allocation concealment were both associated with inflated effect sizes in an analysis of 250 trials from 33 meta-analytic reviews. Thus, we believe that limitations in the quality of trial reporting are likely to provide a reasonable proxy of trial quality.

Study results must also be interpreted in the context of the review objectives. Because the review aimed to assess the quality of evidence in support of different approaches and not the quantity of evidence, conclusions about the relative effectiveness of specific interventions are limited. An approach supported by three methodologically stronger studies is supported by higher quality evidence than is an approach supported by two stronger studies, but it is not necessarily a superior model. Quantification of evidence in support of different interventions by consideration of effect sizes instead of null significance testing as well as testing whether the attributes identified in this review protect against inflated effect estimates represent important avenues for future research. Furthermore, the focus on internal validity limits conclusions about the external validity, durability, or specificity of outcomes to other domains of adolescent functioning. A number of study attributes included in the CONSORT statement and MQRS, which were not included in this review, such as adverse events, follow-up length, and multisite replication, could be utilized in future reviews to evaluate the reporting of these important dimensions of randomized trials.

Finally, the specific selection criteria necessarily limited the generalizability of the results. The intentional focus on studies published in peer-reviewed journals presumably introduced bias in favor of methodologically stronger studies and precludes inferences about the quality of studies published in other outlets. It also precludes conclusions about studies that may have used high-quality designs but remain unpublished for other reasons, such as negative results. Generalizability is also limited by the identification of only 31 randomized evaluations of outpatient interventions. A 2002 review of the adult literature (Miller & Wilbourne, 2002) focusing specifically on alcohol dependence treatments identified more than 300 studies, a number which stand in stark contrast to the number of randomized trials of adolescent substance abuse treatments identified in the current review.

Research and Clinical Implications

Treatment research in the field of adolescent substance abuse, which is quite challenging to conduct, has nonetheless been characterized by several improvements in quality of evidence over the past 2 decades. Most studies now report baseline data, use treatment manuals, supplement adolescent self-report with collateral reports or objective measures, and compare the treatment of interest to an active condition. At the same time, there are clear methodological attributes that need to be incorporated into the design of future studies if the evidence base in support of outpatient psychological interventions is to improve and to yield more accurate estimates of outcome.

Several of the needed attributes are relatively easy to implement. For example, investigators can readily state their objectives and demarcate primary outcomes. Similarly, procedures to ensure a truly random allocation sequence and to conceal allocation until after baseline assessment are both important and readily available. Other attributes require adjustments in study administration and have associated costs—such as systematically using treatment adherence ratings, assuring that evaluators are kept blind to treatment condition, and retaining treatment dropouts in the assessment schedule—in order to facilitate ITT analyses with minimal reliance on imputed data. The most challenging attribute is likely to be recruiting large enough samples to ensure adequate power when comparing two or more active interventions. At a minimum, investigators can readily describe how sample size was determined and what degree of power is associated with the actual sample size.

Medical reviews have identified three attributes as most critical in assuring that reported effect sizes are not inflated (Brouwers et al., 2005; Juni et al., 1999; Moja et al., 2005): allocation concealment, ITT analyses, and blind assessment of outcome. These methodological attributes protect against bias potentially associated with assignment to treatment, differential dropout from treatment, and evaluation of treatment outcome. Thus, there is a particular need for emphasis on these three attributes in the design of studies.

In addition to the implications for treatment research design and implementation, our review indicates a need for more uniform reporting requirements in the publications generated by outcome studies. Several of the leading journals in clinical psychology and psychiatry have adopted the CONSORT criteria, as previously noted (e.g. Dulcan, 2001; La Greca, 2005). There is a need to broaden this movement to include a wider array of journals as well as to prioritize those attributes that have been shown to protect against biased effect size estimates. General adoption of guidelines such as the CONSORT would have the additional advantage of promoting uniform reporting requirements and stronger methodological quality across related but distinct disciplines. This advantage is particularly important when treatments are derived from different disciplines and compared or combined in outcome studies.

Clinical service implications of this review pertain to the selection of interventions based on the evidence in support of different options. In general, treatments that are supported by a stronger evidence base should be preferred to those with a weaker evidence base by treatment providers, by adolescents and families seeking services, and by agencies funding clinical services. Consistent with the broader adolescent substance use literature, this review finds that family-based therapy, CBT, brief motivational interventions, adolescent group therapy, parent skills interventions, and integrative approaches were all generally associated with treatment gains over time. However, the only specific models that demonstrated evidence of treatment superiority in two or more methodologically stronger studies were ecological family therapy, CBT, and brief motivational interventions, indicating that these approaches are currently supported by higher quality evidence than are alternate

interventions. Results of the current review support more treatment models than did the review of Vaughn and Howard (2004), which identified one ecological family approach (multidimensional family therapy) and CBT as the interventions with the strongest evidence base.

The relatively unfavorable performance of adolescent group therapy, with the exception of group CBT, warrants particular consideration. Non-CBT adolescent group therapy was the only treatment model in this review that demonstrated inferiority in immediate treatment outcome in more than one methodologically stronger study. Group therapy is a frequently used modality in adolescent substance abuse treatment. Etheridge, Smith, Rounds-Bryant, and Hubbard (2001) found that there was a significant shift toward greater use of the group modality in adolescent outpatient drug abuse treatment during the period from the early 1980s to the mid-1990s. In an evaluation of outpatient programs, Hser and colleagues (2001) found that 78% of programs provided group sessions as a significant component of treatment. Given the widespread use of the group modality, our findings suggest that it is advisable for providers to use group interventions with a more substantial evidence base, such as CBT, or to rely more heavily on other modalities with greater evidentiary support.

To attain state-of-the-art interventions for adolescent substance abuse, efficacy studies should prioritize attributes that have been empirically shown to affect effect size estimates. Once efficacy has been demonstrated in high-quality trials, resources can focus on effectiveness models and treatment dissemination. Advocates for treatments that, to date, lack evidence from higher quality studies may wish to consider the attributes designated in this review when designing studies as well as when reporting outcomes relative to treatment models with a more robust evidence base.

References

Adetugbo, K., & Williams, H. (2000). How well are randomized controlled trials reported in the dermatology literature? *Archives of Dermatology, 136,* 381–385.

Altman, D. G. (1996). Better reporting of randomised controlled trials: The CONSORT statement. *British Medical Journal, 313,* 570.

Azrin, N. H., Donohue, B., Besalel, V. A., Kogan, E. S., & Acierno, A. (1994). Youth drug abuse treatment: A controlled outcome study. *Journal of Child & Adolescent Substance Abuse, 3,* 1–16.

Azrin, N. H., Donohue, B., Teichner, G. A., Crum, T., Howell, J., & DeCato, L. A. (2001). A controlled evaluation and description of individual-cognitive problem solving and family-behavior therapies in dually-diagnosed conduct-disordered and substance-dependent youth. *Journal of Child & Adolescent Substance Abuse, 11,* 1–43.

Bailey, K., Baker, A., Webster, R., & Lewin, T. (2004). Pilot randomized controlled trial of a brief alcohol intervention group for adolescents. *Drug and Alcohol Review, 23,* 157–166.

Baskin, T. W., Tierney, S. C., Minami, T., & Wampold, B. E. (2003). Establishing specificity in psychotherapy: A meta-analysis of structural equivalence of placebo controls. *Journal of Consulting and Clinical Psychology, 71,* 973–979.

Brouwers, M. C., Johnston, M. E., Charette, M. L., Hanna, S. E., Jadad, A. R., & Browman, G. P. (2005). Evaluating the role of quality assessment of primary studies in systematic reviews of cancer practice guidelines. *BMC Medical Research Methodology, 5,* 1–9.

Catalano, R., Hawkins, J., Wells, E., Miller, J., & Brewer, D. (1991). Evaluation of the effectiveness of adolescent drug treatment, assessment of risks for relapse, and promising approaches for relapse prevention. *The International Journal of the Addictions, 25,* 1085–1140.

Chalmers, T. C., Smith, H., Blackburn, B., Silverman, B., Schroeder, B., Reitman, D., et al. (1981). A method for assessing the quality of a randomized control trial. *Controlled Clinical Trials, 2,* 31–49.

Clark, D., Kirisci, L., & Tarter, R. (1998). Adolescent versus adult onset and the development of substance use disorders in males. *Drug and Alcohol Dependence, 49,* 115–121.

Conrad, P. J., Stewart, S. H., Comeau, N., & Maclean, A. M. (2006). Efficacy of cognitive-behavioral interventions targeting personality risk factors for youth alcohol misuse. *Journal of Clinical Child & Adolescent Psychology, 35,* 550–563.

Cook, T., & Campbell, D. (1979). *Quasi-experimentation.* Chicago: Rand McNally.

Deas, D., & Thomas, S. E. (2001). An overview of controlled studies of adolescent substance abuse treatment. *The American Journal on Addictions, 10,* 178–189.

Dennis, M., Godley, S. H., Diamond, G., Tims, F. M., Babor, T., Donaldson, J., et al. (2004). The Cannabis Youth Treatment (CYT) study: Main findings from two randomized trials. *Journal of Substance Abuse Treatment, 27,* 197–213.

DerSimonian, R., Charette, L., McPeek, B., & Mosteller, F. (1982). Reporting on methods in clinical trials. *New England Journal of Medicine, 306,* 1332–1337.

Dulcan, M. (2001). Editor's report. *Journal of the American Academy of Child & Adolescent Psychiatry, 41,* 109.

Ellenberg, J. H. (1996). Intent-to-treat analysis versus as-treated analysis. *Drug Information Journal, 30,* 535–544.

Etheridge, R. M., Smith, J. C., Rounds-Bryant, J. L., & Hubbard, R. L. (2001). Drug abuse treatment and comprehensive services for adolescents. *Journal of Adolescent Research, 16,* 563–589.

Friedman, A. S. (1989). Family therapy vs. parent groups: Effects on adolescent drug abusers. *The American Journal of Family Therapy, 17,* 335–348.

Henggeler, S. W., Halliday-Boykins, C. A., Cunningham, P. B., Randall, J., Shapiro, S. B., & Chapman, J. E. (2006). Juvenile drug court: Enhancing outcomes by integrating evidence-based treatments. *Journal of Consulting and Clinical Psychology, 74,* 42–54.

Henggeler, S. W., Pickrel, S. G., & Brondino, M. J. (1999). Multisystemic treatment of substance-abusing and dependent delinquents: Outcomes, treatment fidelity, and transportability. *Mental Health Services Research, 1,* 171–184.

Hill, A. B. (1961). *Principles of medical statistics.* New York: Oxford University Press.

Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *British Medical Journal, 319,* 670–674.

Howard, K., Krause, M., & Orlinsky, D. (1986). The attrition dilemma: Toward a new strategy for psychotherapy research. *Journal of Consulting and Clinical Psychology, 54,* 106–110.

Hser, Y., Grella, C. E., Hubbard, R. L., Hsieh, S., Fletcher, B. W., Brown, B. S., & Anglin, M. D. (2001). An evaluation of drug treatments for adolescents in 4 US cities. *Archives of General Psychiatry, 58,* 689–695.

Huston, P., & Hoey, J. (1996). CMAJ endorses the CONSORT statement: CONsolidation of standards for reporting trials. *CMAJ, 155,* 1277–1282.

Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, J. M., Gavaghan, D. J., & McQuay, H. J. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials, 17,* 1–12.

Joanning, H., Quinn, W., Thomas, F., & Mullen, R. (1992). Treating adolescent drug abuse: A comparison of family systems therapy, group therapy, and family drug education. *Journal of Marital and Family Therapy, 18,* 345–356.

Juni, P., Altman, D. G., & Egger, M. (2001). Systematic reviews in health

care: Assessing the quality of controlled clinical trials. *British Medical Journal, 323,* 42–46.

Juni, P., Witschi, A., Bloch, R., & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association, 282,* 1054–1060.

Kaminer, Y., Burleson, J., Blitz, C., Sussman, J., & Rounsaville, B. (1998). Psychotherapies for adolescent substance abusers: A pilot study. *The Journal of Nervous and Mental Disease, 186,* 684–690.

Kaminer, Y., Burleson, J. A., & Goldberger, R. (2002). Cognitive-behavioral coping skills and psychoeducation therapies for adolescent substance abuse. *The Journal of Nervous and Mental Disease, 190,* 737–745.

Kassirer, J. P., & Campion, E. W. (1994). Peer review: Crude and understudied, but indispensable. *Journal of the American Medical Association, 272,* 96–97.

Kazdin, A. E., & Bass, D. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. *Journal of Consulting and Clinical Psychology, 57,* 138–147.

Kazdin, A., Bass, D., Ayers, W., & Rodgers, A. (1990). Empirical and clinical focus of child and adolescent psychotherapy research. *Journal of Consulting and Clinical Psychology, 58,* 729–740.

Kjaergard, L. L., Villumsen, J., & Gluud, C. (2001). Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Annals of Internal Medicine, 135,* 982–989.

La Greca, A. M. (2005). Editorial. *Journal of Consulting and Clinical Psychology, 73,* 3–5.

Landis, J. R., & Koch, G. G. (1977). A one-way components of variance model for categorical data. *Biometrics, 33,* 671–679.

Latimer, W. W., Winters, K. C., D'Zurilla, T., & Nichols, M. P. (2003). Integrated family and cognitive-behavioral therapy for adolescent substance abusers: A Stage I efficacy study. *Drug and Alcohol Dependence, 71,* 303–317.

Lewis, R. A., Piercy, F. P., Sprenkle, D. H., & Trepper, T. S. (1990). Family-based interventions for helping drug-abusing adolescents. *Journal of Adolescent Research, 5,* 82–95.

Liberati, A., Himel, H., & Chalmers, T. (1986). A quality assessment of randomized control trials of primary treatment of breast cancer. *Journal of Clinical Oncology, 4,* 942–951.

Liddle, H. A., Dakof, G. A., Parker, K., Diamond, G. S., Barrett, K., & Tejeda, M. (2001). Multidimensional family therapy for adolescent drug abuse: Results of a randomized clinical trial. *American Journal of Drug and Alcohol Abuse, 27,* 651–688.

Liddle, H. A., Rowe, C. L., Dakof, G. A., Ungaro, R. A., & Henderson, C. E. (2004). Early intervention for adolescent substance abuse: Pretreatment to posttreatment outcomes of a randomized clinical trial comparing multidimensional family therapy and peer group treatment. *Journal of Psychoactive Drugs, 36,* 49–63.

Lonigan, C., Elbert, J., & Johnson, S. (1998). Empirically supported psychosocial interventions for children: An overview. *Journal of Clinical Child Psychology, 27,* 138–145.

Luborsky, L., McLellan, A. T., Diguer, L., Woody, G., & Seligman, D. (1997). The psychotherapist matters: Comparison of outcomes across twenty-two therapists and seven patient samples. *Clinical Psychology: Science and Practice, 4,* 53–65.

Marsden, J., Stillwell, G., Barlow, H., Boys, A., Taylor, C., Hunt, N., et al. (2006). An evaluation of a brief motivational intervention among young ecstasy and cocaine users: No effect on substance and alcohol use outcomes. *Addiction, 101,* 1014–1026.

McCambridge, J., & Strang, J. (2004). The efficacy of single-session motivational interviewing in reducing drug consumption and perceptions of drug-related risk and harm among young people: Results from a multi-site cluster randomized trial. *Addiction, 99,* 39–52.

McGillicuddy, N. B., Rychtarik, R. G., Duquette, J. A., & Morsheimer, E. T. (2001). Development of a skill training program for parents of substance-abusing adolescents. *Journal of Substance Abuse Treatment, 20,* 59–68.

Miller, W., Brown, J., Simpson, T., Handmaker, N., Bien, T., Luckie, L., et al. (1995). What works? A methodological analysis of the alcohol treatment outcome literature. In R. Hester & W. Miller (Eds.), *Handbook of alcoholism treatment approaches: Effective alternatives* (pp. 12–44). Needham Heights, MA: Allyn & Bacon.

Miller, W. R., & Wilbourne, P. L. (2002). Mesa Grande: A methodological analysis of clinical trials of treatment for alcohol use disorders. *Addiction, 97,* 265–277.

Moher, D., Schulz, K. F., & Altman, D. (2001). The CONSORT Statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *Journal of the American Medical Association, 285,* 1987–1991.

Moja, L. P., Telaro, E., D'Amico, R., Moschetti, I., Coe, L., Liberati, A., & the Metaquality Study Group. (2005). Assessment of the methodological quality of primary studies by systematic reviews: Results of the metaquality cross sectional study. *British Medical Journal, 330,* 1–5.

Moyer, A., & Finney, J. W. (2005). Rating methodological quality: Toward improved assessment and investigation. *Accountability in Research, 12,* 229–313.

Muck, R., Zempolich, K. A., Titus, J. C., Fishman, M., Godley, M. D., & Schwebel, R. (2001). An overview of the effectiveness of adolescent substance abuse treatment models. *Youth and Society, 33,* 143–168.

Najavits, L. M., Gallop, R. J., & Weiss, R. D. (2006). Seeking safety therapy for adolescent girls with PTSD and substance use disorders: A randomized controlled trial. *Journal of Behavioral Health Services & Research,* 453–463.

Ozechowski, T. J., & Liddle, H. A. (2000). Family-based therapy for adolescent drug abuse: Knowns and unknowns. *Clinical Child and Family Psychology Review, 3,* 269–298.

Peterson, P. L., Baer, J. S., Wells, E. A., Ginzler, J. A., & Garrett, S. B. (2006). Short-term effects of a brief motivational intervention to reduce alcohol and drug risk among homeless adolescents. *Psychology of Addictive Behaviors, 20,* 254–264.

Plint, A., Moher, D., Morrison, A., Schulz, K., Altman, D., Hill, C., et al. (2006). Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Medical Journal of Australia, 185,* 263–267.

Santisteban, D. A., Coatsworth, J. D., Perez-Vidal, A., Kurtines, W. M., Schwartz, S. J., LaPerriere, A., & Szapocznik, J. (2003). Efficacy of brief strategic family therapy in modifying Hispanic adolescent behavior problems and substance use. *Journal of Family Psychology, 17,* 121–133.

Schulz, K., Chalmers, I., Grimes, D., & Altman, D. (1994). Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *Journal of the American Medical Association, 272,* 125–128.

Schulz, K. F., Chalmers, I., Hayes, R. J., & Altman, D. G. (1995). Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association, 279,* 1489–1491.

Schulz, K. F., & Grimes, D. A. (2002). Generation of allocation sequences in randomised trials: Chance, not choice. *The Lancet, 359,* 515–519.

Slesnick, N., & Prestopnik, J. L. (2005). Ecologically based family therapy outcome with substance abusing runaway adolescents. *Journal of Adolescence, 28,* 277–298.

Smith, D. C., Hall, J. A., Williams, J. K., An, H., & Gotman, N. (2006). Comparative efficacy of family and group treatment for adolescent substance abuse. *The American Journal on Addictions, 15,* 131–136.

Srisurapanont, M., Sombatmai, S., & Boripuntakul, T. (2007). Brief intervention for students with methamphetamine use disorders: A randomized controlled trial. *The American Journal on Addictions, 16,* 111–116.

Stanton, M. D., & Shadish, W. R. (1997). Outcome, attrition, and family-

couples treatment for drug abuse: A meta-analysis and review of the controlled, comparative studies. *Psychological Bulletin, 122,* 170–191.

Substance Abuse and Mental Health Services Administration. (2004). *Treatment Episode Data Set (TEDS). Highlights—2004. National Admissions to Substance Abuse Treatment Services* (DASIS Series S-31, DHHS Publication No. SMA 06–4140). Rockville, MD: Office of Applied Studies.

Substance Abuse and Mental Health Services Administration. (2007). *Results from the 2006 National Survey on Drug Use and Health: National Findings* (NSDUH Series H-28, DHHS Publication No. SMA 05–4062). Rockville, MD: Office of Applied Studies.

Szapocznik, J., Kurtines, W., Foote, F., Perez-Vidal, A., & Hervis, O. (1983). Conjoint versus one-person family therapy: Some evidence for the effectiveness of conducting family therapy through one person. *Journal of Consulting and Clinical Psychology, 51,* 889–899.

Szapocznik, J., Kurtines, W., Foote, F., Perez-Vidal, A., & Hervis, O. (1986). Conjoint versus one-person family therapy: Further evidence for the effectiveness of conducting family therapy through one person with drug-abusing adolescents. *Journal of Consulting and Clinical Psychology, 54,* 395–397.

Thornley, B., & Adams, C. (1998). Content and quality of 2,000 controlled trials in schizophrenia over 50 years. *British Medical Journal, 317,* 1181–1184.

Thush, C., Wiers, R. W., Theunissen, N., Van den Bosch, J., Opdenacker, J., van Empelen, P., et al. (2006). A randomized clinical trial of a targeted intervention to moderate alcohol use and alcohol-related problems in at-risk adolescents. *Pharmacology Biochemistry and Behavior, 86,* 368–376.

Tukey, J. W. (1977, November). Some thoughts on clinical trials, especially problems of multiplicity. *Science, 198,* 679–684.

Vaughn, M. G., & Howard, M. O. (2004). Adolescent substance abuse treatment: A synthesis of controlled evaluations. *Research on Social Work Practice, 14,* 325–335.

Waldron, H. B. (1997). Adolescent substance abuse and family therapy outcome: A review of randomized trials. In T. H. Ollendick & R. J. Print (Eds.), *Advances in clinical child psychology* (Vol. 19, pp. 199–234). New York: Plenum Press.

Waldron, H. B., & Kaminer, Y. (2004). On the learning curve: The emerging evidence supporting cognitive-behavioral therapies for adolescent substance abuse. *Addiction, 99,* 93–105.

Waldron, H. B., Slesnick, N., Brody, J. L., Turner, C. W., & Peterson, T. R. (2001). Treatment outcomes for adolescent substance abuse at 4- and 7-month assessments. *Journal of Consulting and Clinical Psychology, 69,* 802–813.

Walker, D. D., Roffman, R. A., Stephens, R. S., Berghuis, J., & Kim, W. (2006). Motivational enhancement therapy for adolescent marijuana users: A preliminary randomized controlled trial. *Journal of Consulting and Clinical Psychology, 74,* 628–632.

Waltz, J., Addis, M., Koerner, K., & Jacobson, N. (1993). Testing the integrity of the psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61,* 620–630.

Weisz, J. R., McCarty, C. A., & Valeri, S. M. (2006). Effects of psychotherapy for depression in children and adolescents: A meta-analysis. *Psychological Bulletin, 132,* 132–149.

Williams, R. J., Chang, S. Y., & Addiction Centre Adolescent Research Group. (2000). A comprehensive and comparative review of adolescent substance abuse treatment outcome. *American Psychological Association, 7,* 138–166.

Windle, M., & Davies, P. T. (1999). Depression and heavy alcohol use among adolescents: Concurrent and prospective relations. *Development and Psychopathology, 11,* 823–844.

Winters, K. C. (1999). Treating adolescents with substance use disorders: An overview of practice issues and treatment outcome. *Substance Abuse, 20,* 203–224.

Winters, K. C., & Leitten, W. (2007). Brief intervention for drug-abusing adolescents in a school setting. *Psychology of Addictive Behaviors, 21,* 249–254.

Received June 22, 2007
Revision received April 7, 2008
Accepted April 9, 2008 ■