HHS Public Access

Author manuscript

J Fam Psychol. Author manuscript; available in PMC 2015 December 08.

Published in final edited form as:

J Fam Psychol. 2015 December ; 29(6): 919–930. doi:10.1037/fam0000127.

Gender and Ethnicity as Moderators: Integrative Data Analysis of Multidimensional Family Therapy Randomized Clinical Trials

Paul E. Greenbaum,
Department of Child and Family Studies, College of Behavioral and Community Science, University of South Florida

Wei Wang,
Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida

Kristin Hall,
Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida

Craig E. Henderson,
Department of Psychology, Sam Houston State University

Lisa Kan,
Department of Psychology, Sam Houston State University

Gayle A. Dakof, and
Department of Public Health Sciences, University of Miami Miller School of Medicine.

Howard A. Liddle
Department of Public Health Sciences, University of Miami Miller School of Medicine.

Abstract

This study examined gender and ethnicity as moderators of Multidimensional Family Therapy (MDFT) effectiveness for adolescent drug abuse and illustrated the utility of integrative data analysis (IDA, Bauer & Hussong, 2009) for assessing moderation. By pooling participant data from five independent MDFT randomized clinical trials (RCTs), IDA increased power to test moderation. Participants were 646 adolescents receiving treatment for drug use, aged 11 to 17 years (M = 15.31, SD = 1.30), with 19% female (n = 126), 14% (n = 92) European American, 35% (n = 225) Hispanic, and 51% (n = 329) African American. Participants were randomized to MDFT or active comparison treatments, which varied by study. Drug use involvement (i.e., frequency and consequences) was measured at study entry, 6-, and 12-months by a four-indicator latent variable. Growth curve change parameters from multiple calibration samples were regressed on treatment effects overall and by moderator subgroups. MDFT reduced drug use involvement (p < .05) for all

Correspondence concerning this article should be addressed to Paul E. Greenbaum, 517 Montrose Avenue, Temple Terrace, FL 3316. greenbap@usf.edu..

Disclosure: Howard A. Liddle receives financial compensation for his role as consultant and member of the Board of Directors of MDFT International, a 501(c) (3) public charity dedicated to the implementation of MDFT. Gayle A. Dakof receives financial compensation for her role as the Director of MDFT International.

participant groups. Pooled comparison groups reduced drug use involvement only for females and Hispanics (*p*s < .05). MDFT was more effective than comparisons for males, African Americans, and European Americans (*p*s <.05; Cohen's *d* = 1.17, 1.95, and 1.75, respectively). For females and Hispanics, there were no significant differences between MDFT and pooled comparison treatments, Cohen's *d* = 0.63 and 0.19, respectively. MDFT is an effective treatment for drug use among adolescents of both genders and varied ethnicity with males, African American, and White Non-Hispanic adolescents benefitting most from MDFT.

Keywords

adolescents; integrative data analysis; drugs; treatment

The extent that participant characteristics are associated with outcome disparities for behavioral treatments generally, and adolescent drug abuse treatments in particular, is arguably one of the most important, albeit difficult to answer questions in the psychotherapy field (Blake, Amaro, Schwartz, & Flinchbaugh, 2001). Limited knowledge about gender and ethnicity as moderators remains ubiquitous in treatment research. Sparse representation of minority individuals in RCTs and paucity of subgroup analyses based on ethnicity has been well-documented. For example, a recent review of NIMH-funded randomized clinical trials, revealed that only 10% and 5% reported moderation analyses for gender and ethnicity, respectively (Mak, Law, Alvidrez, & Pérez-Stable, 2007). A major impediment for conducting these subgroup analyses has been small to moderate sample sizes (80% had less than 200 participants). Ethnic and gender subgroup sizes have been seldom more than 40 with consequently little power to test moderation (Mak et al., 2007).

From this perspective, the state of the literature stands at odds with a growing emphasis on disseminating evidence-based treatments (EBT) on a wide-scale, and researchers have argued that results from existing RCTs cannot be generalized to individuals from ethnic minority backgrounds (Hall, 2001; Sue, Zane, Hall, & Berger, 2009). Correspondingly, there has been recognition that treatment needs, processes, and outcomes might differ across gender (American Psychological Association, 2007; Blake et al., 2001). Although girls and women generally have been adequately represented in RCTs, few researchers have investigated potential gender differences in treatment effects (Mak et al., 2007).

Gender and Ethnicity Differences in Drug Abuse Treatment Effects

Within research on adolescent substance abuse treatment, recommendations to improve understanding of ethnic and gender differences in preventing and treating drug use have had limited success (Hall, 2001). Few studies have considered participant ethnicity in their study design and even fewer have included subgroup analyses (Strada, Donohue, & Lefforge, 2006). Szapocznik and colleagues (Szapocznik, Prado, Burlew, Williams, & Santisteban, 2007) noted the disappointing number of published studies regarding substance use prevention and treatment for African American and Latino youth, while Blake et al. (2001) and Kumpfer, Smith, and Summerhays (2008) cautioned that existing literature has not established strong empirical support that prevention programs have been equally effective for girls and boys. Treatment studies typically have been designed primarily to test treatment

main effects leaving statistical tests involving subgroups almost always underpowered. Integrative data analysis (IDA) has the potential to maximize existing power based on an augmented *N* provided by pooled data.

The present study addressed this gap in the research literature by conducting an IDA on five RCTs (i.e., Dakof et al., 2015; Liddle, 2008; Liddle et al., 2013; Liddle et al., 2009; Liddle et al., 2011) of Multidimensional Family Therapy (MDFT; Liddle, 2002), an evidenced-based adolescent drug abuse treatment program. Based on Bauer & Hussong's (2009) moderated nonlinear factor analysis approach, IDA enabled pooling of individual study data and provided enhanced power to test moderator effects. Participants in these trials were male and female adolescents of European American, African American, and Hispanic ethnicity, 11 to 17 years of age, with documented drug use (in most cases with a drug use disorder diagnosis) and their families. The number of participants meeting study criteria in each of the RCTs ranged from 83 to 224. In addition to providing a robust test of the MDFT intervention main effect across studies, this study sought to extend previous MDFT research in two ways. First, by pooling individual data, power to detect moderator effects was increased. Previously, those individual MDFT trials that tested subgroup effects had not found gender and ethnicity to be moderators (e.g., Liddle et al., 2008, 2009), but the likelihood of finding significant effects was severely compromised due to low statistical power in the individual studies. Secondly, by modeling drug use involvement as a latent variable, measurement error was attenuated and various observed indicators of drug use involvement across- and within-individual studies were linked.

Multidimensional family therapy for adolescent drug use

MDFT (Liddle et al., 2011) is a developmentally-oriented, family-based, integrative outpatient treatment that has blended family therapy, individual therapy, drug counseling, and multiple-systems oriented interventions to impact important psychosocial systems of adolescents and their families. It is a flexible treatment system designed to be appropriate for various patient populations and client severity levels, with different intensities ranging from prevention and early intervention (Study 2, Liddle et al., 2009) to adolescents deeply involved with drugs and the juvenile justice system (Study 3, Liddle et al., 2013). Results from individual RCTs have indicated that receiving MDFT decreased drug use involvement and increased prosocial behaviors more than comparison treatments (Liddle et al., 2001, 2008, 2009, 2011; Rigter et al., 2012). MDFT also has been associated with greater decreased delinquent behavior (Dakof et al., 2015; Liddle et al., 2009, 2011)), externalizing behaviors (Liddle et al., 2001), and internalizing symptoms (Liddle et al., 2009). In the five studies included in the current IDA, individual study results indicated that MDFT decreased drug use dependence symptoms to a greater extent than individual cognitive behavioral therapy (Liddle et al., 2008) and adolescent group therapy (Liddle et al., 2009) among community adolescents referred for drug abuse treatment and adolescents enrolled in a post-adjudication juvenile drug court (Dakof et al., 2015). MDFT also decreased frequency of drug use to a greater extent than comparison treatments among adolescents that: (a) received adolescent group therapy (Liddle et al., 2009), (b) engaged in treatment while incarcerated in a juvenile detention facility and remained in treatment following detention discharge (Liddle et al., 2011), and (c) referred for residential treatment (Liddle et al., 2013).

MDFT has been implemented in diverse community settings across the United States and internationally, with study samples of male and female adolescents from varied ethnic groups. Study designs have followed CONSORT guidelines, used intent-to-treat analyses, and participated in multisite RCTs involving independent investigators. Process studies have supported theoretical propositions about the model's family-based mechanisms of action (Henderson et al., 2009). At the same time, little is known about how well MDFT works comparatively with males and females and adolescents from different ethnic backgrounds limiting generalizability, as well as the clinical usefulness, of these findings.

Integrative Data Analysis

Integrative data analysis (IDA) is the statistical analysis of two or more independent studies that have been pooled into one (Curran & Hussong, 2009). Additionally, adjustments for measurement noninvariance that include heterogeneity of measurement within- and across-studies and variation in scale type can be accommodated using latent variable modeling approaches. In analyzing moderation, IDA also has advantages over more traditional pooled data techniques. Compared to conventional meta-analysis, which combines summary statistics, IDA, by using individual participant's data, has increased power to detect group effects and can assess participant-level characteristics as moderators without committing the ecological fallacy of attributing study-level effects to individuals (Robinson, 1950). Currently, IDA applications have been limited to studies of developmental psychopathology (Bauer & Hussong, 2009), and, more recently, a substance abuse study investigating measurement equivalence of nicotine dependence symptoms (Rose, Dierker, Hedeker, & Mermelstein, 2013). No studies have used IDA to test moderator treatment effects from RCTs.

The most comprehensive approach for conducting IDA has used moderated nonlinear factor analysis (MNLFA; Bauer & Hussong, 2009) to model a latent variable with commensurate indicators across multiple studies. For the current study, MNLFA performed three important tasks. First, MNLFA accommodated observed indicators with a variety of distributional properties (i.e., continuous, binary, negative binomial, censored normal). Secondly, a commensurate measurement model was achieved using latent variable modeling for indicators not measured, by design, across studies under the missing at random assumption (MAR; Little & Rubin, 2002). Additional adjustments for lack of measurement invariance or differential item functioning (DIF) across studies and other important subgroups (e.g., treatment condition, gender, ethnicity) also were included. Thirdly, non-independence of observations caused by repeated measures was accommodated using a two-stage procedure whereby a randomly-selected calibration sample of independent observations, with a single observation per participant, generated individual factor scores for all observations from all participants. Factor scores were then subjected to latent growth curve modeling. A two-stage approach was necessary because a simultaneous model solution was intractable as existing computer capacity and software cannot accommodate both MNLFA model complexity (nonnormal data distributions, latent mean and variance regression, and DIF testing) and dependency among observations.

The current study conducted an IDA on five RCTs that tested MDFT versus active comparison treatments using multiple indicators of drug use involvement and examining gender and ethnicity as moderators of treatment effects. Specific research questions were:

1. Is MDFT more effective than comparison treatments for males and females?
2. Is MDFT more effective than comparison treatments for African American, Hispanic, and European Americans?

Methods

Sample

Data for this study came from five RCTs that tested effects of MDFT among adolescents receiving drug abuse treatment (Dakof et al. 2015; Liddle, 2008; Liddle et al., 2013; Liddle et al., 2009; Liddle et al., 2011). Among a total of eight MDFT RTCs that have been conducted and had outcome data available for the current study, these five were chosen because of compatibility of participant's ethnicity and outcome measures across studies. One study was excluded because it was conducted with a European multiple country sample with incommensurate ethnicity categories, i.e., no African American, Hispanic data. The other two studies were excluded due to measurement issues. In Liddle et al. (2001), the measure itself—a 10-point clinician rating was substantially different from those used in the included studies. In Dennis et al. (2004), only a single indicator was available for linking with the other studies precluding any study-wide adjustments for measurement invariance in the outcome measure.

All selected studies were conducted in the Unites States. In the pooled sample, participants were randomly assigned to MDFT or one of the active comparison treatments. Study inclusion criteria were determined by the parent projects, which were roughly similar across studies with participants: (a) aged between 11 and 17 years, (b) meeting criteria for a drug disorder or were referred for drug use treatment by an institution, (c) having at least one parent to participate in treatment and research assessments, and (d) providing informed consent/assent to participate in the study. Table 1 lists sample characteristics for each of the individual trials. Participants were 646 adolescents aged 11 to 17 years (M = 15.31, SD = 1.30), with 19% female (n = 126), 14% (n = 92) European American, 35% (n = 225) Hispanic and 51% (n = 329) African American. Participants were primarily cannabis users with a minority of participants also reporting alcohol and other drug use, with cocaine and anxiolytics being the predominant illicit drugs outside of cannabis. Results from an independently-conducted meta-analysis indicated that the effect sizes of the trials we included adequately represented the population of MDFT effects, as the current study included studies producing the largest (Liddle et al., 2009) and among the smallest (Dakof et al., 2015) effect sizes for substance use.

Individual RCTs

The following provides key details on the individual studies comprising the IDA sample. Please see manuscripts for the parent studies and Table 1 for more information. Study 1 (Liddle et al., 2008) was a RTC (n = 224) comparing MDFT and individually-delivered

cognitive behavioral therapy (CBT). Both treatments were delivered: (a) by equally experienced therapists, (b) in equal length and intensity, and (c) in the same format. Participants were between 12 and 17.5 years of age and were recruited primarily from juvenile justice (48%) or child welfare (36%) agencies. Study youth were primarily male and African American, and all were drug users, with 75% meeting criteria for cannabis dependence and 13% meeting criteria for cannabis abuse. Study 2 (Liddle et al., 2009, n = 83) was a randomized effectiveness study in which, by design, MDFT was compared to adolescent group treatment (AGT) among adolescents 15 years or younger. Referrals came from primarily juvenile justice (45%), or schools (41%). Study youth were primarily male and ethnically diverse: 42% Hispanic, 38% African American, 11% Haitian or Jamaican, 3% European American, and 4% other. At intake, 47% met criteria for substance abuse, and 16% met criteria for substance dependence. Study 3 (Liddle et al., 2013) compared an intensive outpatient version of MDFT to inpatient treatment among youth meeting local (Miami) criteria for residential placement. Referrals came from primarily juvenile justice (18%), or a drug abuse assessment and stabilization facility that received most of its referrals from juvenile justice or child welfare systems (67%). Participants were primarily male and Hispanic. All participants met criteria for a drug use disorder, and, by study design, met criteria for at least one comorbid disorder. Study 4 (Liddle et al., 2011) was a detention-to-community study in which participants were enrolled in detention. Participants again were primarily male; most were African American. Approximately 33% of the participants met criteria for cannabis dependence and 29% met criteria for cannabis abuse. Comorbidity was common with participants averaging over two diagnoses at intake. Study 5 was conducted with participants enrolled in a post-adjudication juvenile drug court (Dakof et al. 2015). Participants were primarily male and of either Hispanic or African American ethnicities. At intake, 61% of youth had a cannabis abuse disorder and 30% met criteria for cannabis dependence. Comorbidity was again common, with conduct disorder (52%) being most prevalent.

Procedures

Research procedures for the individual studies were similar. All studies comprising the IDA sample were approved and monitored by Institutional Review Boards (IRBs) at the universities that conducted the studies. Trained research assistants contacted parents and youth, described study purpose and procedures, including randomization, and obtained written informed consent prior to the first assessment session. After baseline assessment, adolescents were randomly assigned to either MDFT or an active comparison treatment. Across studies, data were collected at 8 time points: baseline (i.e., entry into treatment), and 6 weeks, and 2, 3, 4, 6, 9, and 12 months following baseline. Data collection periods varied between studies. Follow-up rates were good overall, averaging 85% across studies through 12 months but notably lower in Liddle et al., 2008, (Liddle et al., 2008, 57%; 2009, 97%; 2013, 97%; 2011, 98%; Dakof et al. 2015; 89%). See Figure 1 for a Consolidated Standards of Reporting Trials (CONSORT) flow diagram collapsing participants across studies. Aside from the inclusion criteria for the individual studies, additional criteria for the pooled IDA sample included no missing data for either ethnicity or gender and all study participants came from European American, Hispanic, or African American backgrounds. Please see

individual studies (Dakof et al. 2015; Liddle et al., 2008, 2009, 2011, 2013) for more details on procedures for the individual studies.

Treatments

Core interventions were the same across studies and individuals, and MDFT therapists worked to engineer change in four major life domains: (a) adolescent (intrapersonal and development issues), (b) parent(s) (individual functioning of the parent as well as parenting practices), (c) family environment (family interactions), and (d) community systems influencing adolescent and family (e.g., working with schools, social service agencies, and juvenile justice). In Study 1, both MDFT and CBT were office-based and administered once per week (60-90 minute sessions). In the other studies, MDFT was delivered both in a clinic and participants' homes with approximately 1-3 sessions per week over the course of 3 to 6 months of treatment. As reported in the individual studies, an acceptable degree of treatment fidelity to MDFT and comparison treatments was achieved. MDFT fidelity was assessed by trained raters using the Therapist Behavior Rating Scale (TBRS, Hogue et al., 1998) or the MDFT Intervention Inventory (Rowe et al., 2013); both scales are validated instruments with strong psychometric properties. With the exception of Liddle et al. (2008), which also used the TBRS to assess fidelity, other studies used observational checklists completed by trained observers to assess the fidelity of comparison treatments. Please see Rowe et al. (2013) for more information on treatment fidelity for the MDFT interventions. Across studies, MDFT more effectively retained youth in treatment than the comparisons; for MDFT the average treatment episode lasted 4.63 months (*SD* = 3.07), for comparisons, treatment lasted an average of 2.73 months (*SD* = 2.71).

Comparison treatments, which varied between studies, included individual CBT, residential treatment, and AGT. With the exception of residential treatment provided in Study 3, all were office-based. The common thread running through the comparison treatments was a theoretical foundation of CBT and a focus on the individual adolescent, with minimal to no attention paid to either modifying family dynamics or directly influencing external systems. AGT tended to be based on a risk and protective factor framework, seeking to reduce drug use directly and focusing on accompanying risk factors (e.g., antisocial peers, low self-esteem, poor academic performance, limited social skills). Groups were "open" (i.e., new members were admitted on a rolling basis) and had approximately 4-6 adolescents led by a single therapist 2-3 times per week. In Study 1, individual CBT was based on broadly defined cognitive behavioral theory that conceptualized drug use as learned behavior started and maintained in the context of environmental factors. Treatment occurred in three stages: (a) prioritizing problems and developing a treatment contract; (b) implementing CBT to increase coping and reduce injurious behaviors; and (c) preventing relapse. In Study 3, residential treatment was based on a social learning approach that emphasized positive reinforcement for appropriate coping behavior and social skills and followed a schedule whereby adolescents completed therapeutic activities each week. All comparison treatments were manualized, therapists received initial training and ongoing supervision in the treatment model, and fidelity to the model was assessed. However, with the exception of the CBT provided in Study 1, none of the comparison treatments were evidence-based per se, as they had not been tested in previous controlled trials.

Measures

To reduce measurement error, multiple measures of drug use involvement, which included objective and self-reported drug frequency and self-reported drug-related consequences, were conceptualized as representing an underlying drug involvement latent variable. The combination of consumption and associated problems measures as indicators of the outcome latent variable provided a multidimensional set of clinical indicators associated with drug use disorders. The latent construct was measured by four observed variables: (a) a positive urinalysis for any of five substances (i.e., benzodiazapines, cocaine, amphetamine, opiates, and marijuana, (b) 30 day TimeLine Follow Back (TLFB; Sobell & Sobell, 1992), (c) the Personal Involvement with Chemicals (PIC) scale of the Personal Experience Inventory (PEI; Winters & Henly, 1989), and (d) the Problem Oriented Screening Instrument for Teens (POSIT; Rahdert, 1991).

TimeLine Follow Back—The TLFB obtains retrospective reports of daily drug use by employing a calendar and other memory prompts to stimulate recall. Youth reported on specific substances used daily for either a 30-day (Liddle et al., 2008, 2009, 2013) or 90-day period (Dakof et al. 2015; Liddle et al., 2011) prior to each assessment. The 90-day and 30-day time periods were harmonized by only using data for the 30 days preceding the assessment with both 30- and 90-day timelines and calculating a total drug use score corresponding to the number of days participants had used any of the five specified substances in the previous 30 days.

Personal Involvement with Chemicals—The PIC is a 29-item scale focusing on the psychological and behavioral depth of drug use involvement and related consequences in the previous 30 days. Items addressed issues such as using substances to feel calm; using them during the whole day, weekends, or at school; and canceling plans to get high. Widely used in applied research settings, it has demonstrated excellent reliability and validity across samples of adolescents from multiple ethnic backgrounds (Winters, Latimer, Stinchfield, & Egan, 2004). Internal consistency reliabilities for the PIC among the IDA sample ranged from .94 – .95.

Problem Oriented Screening Instrument for Teens—The POSIT is a self-report multi-problem screening instrument designed to screen for substance use and other problems. It is a well-validated and reliable instrument that has been widely used in applied treatment settings. The 17-item Substance Use and Abuse subscale was used in this study. Internal consistency reliabilities for the POSIT for the studies comprising the IDA sample ranged from .80 - .85.

Urinalysis—Urine samples were collected at each assessment to substantiate self-reports. Samples were sent to a toxicology laboratory that analyzed specimens using state-of-the-art screening and confirmation methodologies. Screening was conducted for the following substances: THC, amphetamines, benzodiazepines, cocaine, and opiates. Urine was first screened by an enzyme immunoassay (EIA), and then confirmed and quantified by gas chromatography.

Scaling of measures—The urinalysis was a binary variable and scored as 1 if any of the five substances resulted in a positive test and 0 if all substances were negative. The TLFB and POSIT were count variables analyzed as negative binomial distributions to accommodate overdispersion (i.e., conditional variance exceeds the conditional mean), which was significant ($p < .001$) for both variables. The TLFB was scored from 0 to 150 based on the number of occasions that a participant had used each of the five substances during the 30-day period. The 17-item POSIT was scored from 0 to 17. Scoring of the PEI and POSIT followed procedures discussed in the test manuals. The 29-item PEI was scored from 0 to 29 and analyzed as a censored, from below, normal variable.

Analytic Approach

Overview—Analyses were conducted using software packages Mplus v. 7.00 (Muthén & Muthén, 1998-2012) and SAS v. 9.2. The IDA followed Bauer and Hussong's (2009) MNLFA procedure and the multiple calibration sample extension (Wang et al., 2013). Initially, the four observed measures of drug use involvement were tested using confirmatory factor analyses (CFA) as manifest indicators of a latent variable. After establishing the measurement model, the analysis proceeded using a two-stage approach. In stage one, an initial calibration sample of one observation per participant was randomly drawn from the pooled data. MNLFA was then used to produce latent factor scores of drug use involvement for all observations from the five MDFT RCT studies. In the second stage, maximum a posteriori (MAP) probability factor scores were analyzed for intervention effects using growth curve modeling. The two-stage process was repeated on 20 multiple calibration samples to achieve stable combined parameter estimates.

Confirmatory factor analysis—In the first step of the analysis, based on the total sample, CFAs were conducted at the item level for the two multi-item measures, the PEI and POSIT, to assess if each of the observed summary scale scores represented a common factor. After establishing adequate fit of a one-factor model for each scale, a CFA tested if the four observed indicators (i.e., PEI summary score, POSIT summary score, TLFB score, and binary urine analysis outcome) had significant loadings on a common factor latent variable.

Differential item functioning—Next, using the calibration sample, subgroup main and interaction effects with respect to the latent mean and variance of the drug use latent variable were tested. Tested subgroups included treatment condition (MDFT vs. comparison), study, gender, ethnicity, and all of the two-way treatment interactions (i.e., treatment by study, treatment by gender, treatment by ethnicity). All covariates that were significant at the $p <$ .05-level were retained in the calibration sample's model. Latent mean and variance tests were conducted prior to testing for measurement invariance (or differential item functioning [DIF]) across these subgroups because differences in either the level of the latent mean or latent variance across studies or subgroups could otherwise be mistaken for DIF. DIF was tested for the three indicators that were self-report measures, and indicator intercepts and slopes were adjusted (freed) for any significant ($p > .05$) DIF parameter that existed. DIF was not tested for the urinalysis measure as it was an objective biological measure ostensibly not subject to measurement noninvariance. A final calibration model was

developed that included all latent mean effects, all latent variances with *p*-levels < .10, and all significant ($p < .05$) DIF covariate effects. These levels of significance were chosen to include as many covariate terms as possible while maintaining a balance between model parsimony and goodness of fit. This strategy was followed to mitigate differences between factor score estimates and true scores.

Growth curve analysis—In the next step of the analysis, a final calibration model generated MAP scores (i.e., the mode of the latent factor posterior distribution for person *j*) for all participants at all measurement occasions by creating a 'dummy' indicator and using the SAS macro nlmixed. For further details on this process refer to Supplementary Material for Bauer and Hussong (2009). Quality of the factor scores was assessed by plotting the standard errors of the MAP scores for the calibration sample. Factor scores at intake, and 6 and 12 months following intake were then analyzed using latent growth curve modeling. These assessment points were selected because at each assessment at least four of the five studies contributed data and minimized potential bias that might be caused by a study by time confound. A log time scale was used in the growth modeling as change over time was nonlinear and there were insufficient degrees of freedom to fit a quadratic growth parameter. Subsequently, the unconditional growth model was expanded to include covariates of the growth parameters. The intercept and linear slope growth parameters were regressed on study, treatment, gender, and ethnicity main effects as well as treatment by gender, treatment by study, and treatment by ethnicity interactions. Of particular interest were the tests of the treatment main effect and the moderator treatment effects on the slope parameters representing change during the 12 month study period.

Multiple calibrations—Following Wang et al.'s (2013) multiple calibration method, the preceding growth modeling were repeated on 20 calibration samples drawn with replacement. Multiple calibrations were combined to increase reliability of the final estimates and eliminate potential selection bias introduced by using a single calibration. Final estimates were an average of the individual calibrations. The standard error of the final estimates incorporated both within- and between-calibration sample variability. In theory, an estimate based on a large number of calibration samples will eventually converge to the maximum likelihood estimate (MLE), however; in practice, obtaining the number of samples to reach convergence would be very time consuming and may only provide marginal gain with respect to precision and statistical power over a pre-specified number of samples. A simulation study designed to assess the optimal number of calibrations found that 20 calibrations often achieved stability and provided satisfactory results in comparison to MLE (Wang et al., 2013). Following this guideline, this study similarly relied on final estimates that combined results from 20 calibrations.

Results

Confirmatory factor analyses

The POSIT and the PEI, two of the four indicators used to measure drug use involvement, were summary scores derived from scales composed of multiple items. An underlying assumption of the summary scores was that the corresponding-scale represented a common

factor. CFAs tested this assumption. For the 17 true-false items of the POSIT, a single-factor latent variable model using a categorical estimator (WLSMV) and robust standard errors for dependent repeated observations adequately fit the combined sample data, χ^2 (119, N = 646) = 468.37, p <.001, RMSEA = 0.033, CFI = .961, TLI = .955. All item loadings were significant, ps < .001 with standardized loadings ranging from .566 to .819 and an average loading of .731. Similar results were found for the 29-item, four-response categories, PEI. A single-factor latent variable model using a categorical estimator (WLSMV) and robust standard errors for dependent repeated observations fit the data well, χ^2 (594, N = 646) = 3,850.58, p <.001, RMSEA = 0.044, CFI = .955, TLI = .952. All item loadings were significant, ps < .001 with standardized loadings ranging from .414 to .865 and an average loading of .774.

Having established that the POSIT and PEI summary scores were reasonable indicators of unidimensional scales, the four indicators (i.e., urinalysis, TLFB, POSIT, PEI) were tested as a one-factor latent variable of the drug use involvement construct. CFA using robust maximum likelihood estimation and robust standard errors for dependent repeated measures observations indicated that all four indicators loaded positively and significantly on the single factor, ps <.001. Traditional fit indices for assessing goodness-of-fit were not available for this model that included count variables among the indicators. However, the significance of the loadings indicated that as the level of the latent factor increased, the participant had a higher probability of a positive urine test, used drugs more often in the last 30 days, had more involvement with drugs, and experienced more drug-related problems. This pattern was consistent with a latent factor representing drug use involvement. Loadings for the indicators were: PEI, 21.66, se = 0.60, t = 36.19, p < .001; POSIT, 1.28, se = .05, t = 24.21, p < .001; Urinalysis, 1.460, se = 0.10, t = 14.56, p < .001; TLFB, 2.231, se = 0.07, t = 30.09, p < .001.

Latent mean, latent variance, and DIF analyses

Twenty calibration samples were drawn by randomly selecting one observation from each participant's repeated measures using SAS Proc SurveySelect. Covariate analyses of drug use involvement mean and variance and indicator DIF analysis were conducted for each calibration sample. In order for a calibration sample to be included, the final calibration model that included latent mean, latent variance, and DIF effects had to converge to a proper solution.

Among latent factor means for the multiple calibration samples, there was a pattern of significance (i.e., p <.05 in 50% or more of the samples) for higher mean drug use involvement over the entire study period in Liddle et al. (2011) and lower in Liddle et al. (2009) compared to the Dakof et al. (2015) reference group. European American participants also had a pattern of significantly higher means than African American participants. Among the latent variance tests, the follow-up assessments revealed a pattern of significantly greater variances compared to baseline. This pattern of reduced variability at baseline is well known in clinical trials. DIF tests indicated that for the POSIT, there was a pattern of significantly larger intercepts for Hispanic and European American compared to African American participants and smaller intercepts for participants in Liddle et al. (2011)

and Liddle et al. (2013) compared to Dakof et al. (2015). For the TLFB, the only consistent DIF pattern was that for the follow-up conditions there were significantly greater intercepts compared to baseline. None of the other latent mean and variance and DIF tests indicated consistent patterns of significance although among individual calibrations there were a few additional terms that were significant at the .05 *p*-level. These additional terms were retained in the measurement model for that individual calibration.

After completing DIF analyses, for each calibration sample, MAP factor scores were generated for all participants at all of the assessed time points. To evaluate the quality of the factor scores, scatter plots of the estimates as a function of the standard errors of the MAP scores for the calibration samples were produced. Figure 2 displays the plot for calibration 17, which was a typical pattern for the calibrations. The pattern indicated that the standard errors were relatively small for scores within three standard deviations of the latent construct mean and provided little information from MAP scores that were low or high (e.g., <−3.00, >3.00).

Growth curve modeling

Prior to growth curve analysis, the pooled MDFT and comparison groups were tested for differences in relative frequencies of the moderators, gender and ethnicity, and study membership. No significant (p <.05) differences were found for these factors, Gender, χ^2 (1, N = 646) = 0.36, p =.55; Ethnicity, χ^2 (2, N = 646) = 2.73, p = .26; Study, χ^2 (4, N = 646) = 0.13, p = .998. Additionally, cell sizes of the MDFT and Comparisons groups were equal, *n*s = 323.

Factor scores from intake, 6 and 12 month assessments were analyzed using latent growth curve modeling. Unconditional linear slope models were fit to the multiple calibrations. Model fit statistics indicated that the linear log-time model was an adequate to good fit for all 20 calibrations, RMSEA *Mdn* = .040, range = .000-.145; CFI *Mdn* = .999, range = .944-1.000; TLI *Mdn* = .998, range = .833-1.019. The unconditional models were then expanded to include covariates. For the intercept and slope growth parameters (I, S), covariates included main effects for treatment, gender, ethnicity (European American, African American, Hispanic), and study. We considered including age as a covariate but ultimately decided not to due to its restricted range and lack of impact on treatment outcomes in preliminary analyses. Additionally, interaction terms for treatment by gender, ethnicity, and study were included with contrasts that tested the treatment effects within the various levels of gender and ethnicity.

Results of the separate calibration model parameter estimates were combined using Wang et al.'s (2013) procedure to produce a combined mean and a conservative estimate of the standard error of the combined mean estimator by incorporating both within- and between-calibration sample variability, which served as the basis for a pseudo *t*-test. Final combined estimates indicated that at entry into the study, there was no significant difference between MDFT and the comparison groups, Intercept b = −0.017, se = 0.032, t = −0.50, p = .62. Neither did ethnic groups differ significantly in initial level overall nor by treatment condition, *p*s >.05. There was a main effect for gender with females at entry having higher

levels of drug use involvement than males in both treatment conditions, $b = 0.126$, $se = 0.055$, $t = 2.44$, $p < .05$.

There was a significant main effect for Time, Slope $b = -0.215$, $se = 0.077$, $t = -2.72$, $p < .01$ indicating that in both treatment conditions, participants decreased drug use involvement during the 12 month period. A significant Time x Gender interaction indicated that for both treatments, females declined more than males, $b = -0.122$, $se = 0.055$, $t = -2.34$, $p < .05$.

Within-treatment contrasts indicated that for MDFT participants, drug use involvement declined significantly, $b = -0.307$, $se = 0.071$, $t = -4.26$, $p < .001$, but for comparison participants, drug use involvement did not decline significantly, $b = -0.122$, $se = 0.089$, $t = -1.34$, $p = .18$. Additionally, a direct test of the Time by Treatment interaction indicated that MDFT participants had decreased drug use involvement significantly more than the participants in comparison groups, $b = -0.185$, $se = 0.045$, $t = -4.07$, $p < .001$. Cohen's d for the MDFT treatment effect was 1.05, representing a large effect size. Panel A of Figure 3 graphically displays the growth curves from the combined estimates for MDFT and comparisons during the study period. Within this overall trend, a significant Time x MDFT x Hispanic interaction indicated that Hispanics receiving MDFT did not decline as much as African American or European Americans receiving MDFT, $b = 0.312$, $se = 0.114$, $t = 2.74$, $p < .01$.

Moderation

Analyses of moderator (i.e., within-gender and within-ethnicity treatment) effects indicated that for males, European Americans, and African Americans, who received MDFT, drug use involvement declined significantly during the 12 month period, male; slope $b = -0.290$, $se = 0.071$, $t = -4.14$, $p < .001$; European American; slope $b = -0.273$, se = 0.095, $t = -2.85$, p < .01; African American; slope $b = -0.380$, $se = 0.071$, $t = -5.30$, $p < .001$. Similar to the finding for the treatment main effects, among corresponding comparison groups, drug use involvement did not decline significantly, male; slope $b = -0.093$, $se = 0.095$, $t = -1.00$, $p = .32$; European American; slope $b = 0.000$, se = 0.130, $t = 0.002$, $p = .998$; African American; slope $b = -0.083$, $se = 0.100$, $t = -0.81$, $p = .42$. Within each of these participant subgroups, a test of the treatment interaction indicated that the MDFT decline was significantly greater than the corresponding decline in the comparison groups, male; slope $b = -0.197$, $se = 0.055$, $t = -3.89$, $p < .001$; European Americans; slope $b = -0.273$, se = 0.101, t = −2.70, p < .01; African American; slope $b = -1.443$, $se = 0.457$, $t = -3.16$, $p < .001$. Effect sizes for the difference by treatment, measured by Cohen's d, were 1.17, 1.95, and 1.75, respectively. Panels B, E, and F of Figure 3 display the growth curves for these subgroups by treatment.

For female and Hispanic adolescents, results did not differ by treatment. Participants who received either treatment showed significant declines in drug use involvement, female MDFT slope $b = -0.381$, $se = .084$, $t = -4.40$, $p < .001$; female comparisons slope $b = -0.246$, $se = 0.089$, $t = -2.75$, $p < .01$; Hispanic MDFT slope $b = -0.215$, $se = .105$, $t = -2.02$, $p < .05$; Hispanic comparisons slope $b = -0.230$, $se = .084$, $t = -2.66$, $p < .01$. Direct tests of the within-group treatment by slope interactions confirmed that the declines by treatment did not differ significantly, female slope $b = -0.134$, $se = .101$, $t = -1.20$, $p = .23$;

Hispanic slope $b = 0.015$, $se = .063$, $t = 0.24$, $p = .81$. Effect sizes were $d = 0.63$ and 0.19, respectively.

Low power is another factor in assessing the null treatment effects for females and Hispanics. Although a principal advantage of IDA is maximization of existing power, in this study, female and Hispanic subgroups had relatively small *n*s, $n = 126$ and 225, respectively, There was adequate power to detect large treatment effects ($d = 0.90$, Lipsey, 1990), but power to detect medium ($d = 0.45$) effects was less than .80. For females, power was .53, and for Hispanics, .78. All other subgroup comparisons had power greater than .80. Panels C and D of Figure 3 display the growth curves for females and Hispanics by treatment.

Discussion

Results from IDA of the pooled data from five RCTs that compared MDFT versus active comparison treatments indicated that during the 12-month study period, participants who received MDFT showed significant declines in drug use involvement. For the comparison treatments, only female and Hispanic participants showed significant declines, whereas male, African American, and European American participants in the comparison groups, either did not decline or showed nonsignificant declines. Additionally, for male, African American, and European American participants there was a significant incremental treatment effect for MDFT when compared to the comparison treatments. No incremental effects were found for females and Hispanic participants. These results support MDFT as an effective drug abuse treatment with adolescents of both genders and varied ethnicity.

Additionally, results support Huey and Polo's (2008) assertion that MDFT is well-established for working with adolescents from ethnic minority backgrounds. Male, European American, and African American adolescents who received MDFT decreased drug use involvement significantly more than comparison treatment peers. MDFT also was effective for female and Hispanic adolescents, although it was not more effective than active comparison treatments as females and Hispanic adolescents were the only two groups who showed benefit from comparison treatments. Another consideration is low power in this study to detect medium size effects for these groups, suggesting a need for additional research.

Because African Americans and males benefit more from MDFT than comparison treatments, it seems as if African American males would benefit most. The fact that MDFT is achieving these effects is encouraging given African American, male youth are disproportionately represented in the juvenile justice system, are underrepresented in treatment, and frequently drop out of treatment early. A comprehensive treatment intervening in multiple systems of influence in young African Americans' lives, such as MDFT, may be necessary to combat numerous barriers to treatment such youth face. It is not necessarily surprising that MDFT would possess cultural synchrony with this group, as an awareness of culturally meaningful themes for African American youth, and ways to incorporate them in therapy, have been developed within MDFT_(Jackson-Gilfort, Liddle, Tejeda, & Dakof, 2001).

On the other hand, females fared almost as well with comparison treatments as with MDFT, and across treatments, females benefitted the most. It is not entirely clear why females, and especially Hispanic females, were particularly responsive to treatment, and this should be an area for future research. Further, females were the most impaired group at treatment entry, and results from other MDFT studies have suggested that more severely impaired adolescents benefit the most from MDFT (Henderson et al., 2010).

This study was the first to show MDFT's effectiveness when multiple indicators of drug use involvement, measured as a latent variable, represented treatment outcome. By using latent a variable, conclusions regarding MDFT's effectiveness are strengthened in at least two ways. First, the amount of error has been reduced since only shared variance across indicators is represented in the latent outcome. Second, any inadvertent “cherry picking” from studies that have used multiple indicators of drug use involvement, but as is conventional, report only significant results for individual indicators that demonstrate treatment effects, has been eliminated. Another strength of this study was MNLFA's ability to incorporate indicators with varied and non-normal distributions into a latent variable. Typically, measures of drug use involvement have non-normal distributions that have hampered advanced statistical analyses or were analyzed using nonoptimal methods (Atkins, Baldwin, Zheng, Gallop, & Neighbors, 2013). Results support the conclusion that IDA parameter estimates need to be based on multiple calibrations. As originally proposed (Curran & Hussong, 2009), only a single calibration provided factor score estimates for performing IDA, but as found in this study, there is sufficient between-sample variability among estimates derived from multiple calibrations, such that reliance on a single calibration would contribute to unreliable results. By using multiple calibrations and Wang et al.'s (2013) procedure for combining parameter estimates, between-sample variability can be taken into account when calculating the standard errors used in statistical testing and stable final estimates can be obtained.

One limitation of the current study is that all comparison treatments were grouped together. Hence, conclusions about ineffectiveness of specific comparison groups is unwarranted. Further, it cannot be ruled out that longer treatment episodes, obtained in the MDFT condition, are responsible for treatment differences rather than specific interventions. With Hispanic adolescents, idiosyncratic study characteristics also may have contributed to the lack of treatment differences. Future analyses including longer follow-up durations may help clarify these findings. Generalizability of study results is also limited in some degree as similar research teams (and one investigator) were involved in all of the studies.

Another study limitation includes the inability to use all available time points in the growth modeling. Because some assessment times were represented by only a single study, these observations confounded time and study and were not included in the growth analysis. A related limitation arises from the study's inability to model interactions between gender, ethnicity, and study, because of the relatively small numbers in the interaction cells resulting in insufficient power for these tests. Notwithstanding that tests of the *pooled* MDFT and comparison samples did not differ significantly on any of these factors, the uneven distribution of gender and ethnicity by study may have confounded treatment and study effects. Future IDA research aimed would benefit by making every effort to include as many

studies as possible so that the maximum number of subgroup relationships could be assessed and generalizability increased.

Acknowledgments

The work reported was supported by the National Institute on Drug Abuse, Grant R01 DA 029089 (PIs Greenbaum and Henderson). The authors acknowledge the contributions of Anders Alexandersson, Daniel Bauer, C. Hendricks Brown, Patrick Curran, Stanley Huey, and our colleagues in the Prevention Science and Methodology Group for their thoughtful comments and valuable discussions on analyzing pooled data. We would also like to thank Benjamin Klekamp and Crystal Connor for their help in data analysis

References

American Psychological Association. Guidelines for psychological practice with girls and women. American Psychologist. 2007; 62:949–979. doi:10.1037/0003-066X.62.9.949. [PubMed: 18085843]

Atkins DC, Baldwin SA, Zheng C, Gallop RJ, Neighbors C. A tutorial on count regression and zero-altered count models for longitudinal substance use data. Psychology of Addictive Behaviors. 2013; 27:166–177. doi:10.1037/a0029508. [PubMed: 22905895]

Bauer DJ, Hussong AM. Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. Psychological Methods. 2009; 14:101–125. doi: 10.1037/a0015583. [PubMed: 19485624]

Blake SM, Amaro H, Schwartz PM, Flinchbaugh LJ. A Review of substance abuse prevention interventions for young adolescent girls. The Journal of Early Adolescence. 2001; 21:294–324. doi: 10.1177/0272431601021003003.

Curran PJ, Hussong AM. Integrative data analysis: The simultaneous analysis of multiple data sets. Psychological Methods. 2009; 14:81–100. doi:10.1037/a0015914. [PubMed: 19485623]

Dakof G,A, Henderson CE, Rowe CL, Boustani M, Greenbaum PE, Wang W, Hawes S, Linares C, Liddle HA. A randomized controlled trial of Multidimensional Family Therapy in juvenile drug court. Journal of Family Psychology. 2015; 29:232–241. doi:10.1037/fam0000053. [PubMed: 25621927]

Godley SH, Hedges K, Hunter B. Gender and racial differences in treatment process and outcome among participants in the adolescent community reinforcement approach. Psychology of Addictive Behaviors. 2011; 25:143–154. doi:10.1037/a0022179. [PubMed: 21443309]

Hall GCN. Psychotherapy research with ethnic minorities: Empirical, ethical, and conceptual issues. Journal of Consulting and Clinical Psychology. 2001; 69:502–510. doi: 10.1037//0022-006X.69.3.502. [PubMed: 11495179]

Henderson CE, Rowe CL, Dakof GA, Hawes SW, Liddle HA. Parenting practices as mediators of treatment effects in an early-intervention trial of multidimensional family therapy. American Journal of Drug and Alcohol Abuse. 2009; 35:220–226. doi:10.1080/00952990903005890. [PubMed: 20180674]

Henderson CE, Dakof GA, Greenbaum P, Liddle HA. Effectiveness of Multidimensional Family Therapy with Higher-Severity Substance Abusing Adolescents: Report from Two Randomized Controlled Trials. Journal of Consulting and Clinical Psychology. 2010; 78:885–897. doi:10.1037/a0020620. [PubMed: 20873891]

Hogue A, Liddle HA, Rowe C, Turner RM, Dakof GA, LaPann K. Treatment adherence and differentiation in individual versus family therapy for adolescent substance abuse. Journal of Counseling Psychology. 1998; 98:104–114. doi:10.1037/0022-0167.45.1.104.

Huey SJ, Polo AJ. Evidence-based psychosocial treatments for ethnic minority youth. Journal of Clinical Child & Adolescent Psychology. 2008; 37:262–301. doi:10.1080/15374410701820174. [PubMed: 18444061]

Hussong AM, Curran PJ, Bauer DJ. Integrative data analysis in clinical psychological research. Annual Review of Clinical Psychology. 2013; 9:61–89. doi:10.1146/annurev-clinpsy-050212185522.

Jackson-Gilfort A, Liddle HA, Tejeda MJ, Dakof GA. Facilitating engagement of African American male adolescents in family therapy: A cultural theme process study. Journal of Black Psychology. 2001; 27:321–340. doi:10.1177/0095798401027003005.

Kumpfer KL, Smith P, Summerhays JF. A wakeup call to the prevention field: Are prevention programs for substance use effective for girls? Substance Use & Misuse. 2008; 43:978–1001. doi:10.1080/10826080801914261. [PubMed: 18649225]

Liddle HA, Dakof GA, Parker K, Diamond GS, Barrett K, Tejeda M. Multidimensional family therapy for adolescent drug abuse: Results of a randomized clinical trial. American Journal of Drug Abuse. 2001; 27:651–688. doi:10.1081/ADA-100107661.

Liddle HA, Rowe CL, Quille TJ, Dakof GA, Mills DS, Sakran E, Biaggi H. Transporting a research-based adolescent drug treatment into practice. Journal of Substance Abuse Treatment. 2002; 22(4): 231–243. doi:10.1016/S0740-5472(02)00239-8. [PubMed: 12072167]

Liddle HA, Dakof GA, Turner RM, Henderson CE, Greenbaum PE. Treating adolescent drug abuse: A randomized trial comparing multidimensional family therapy and cognitive behavior therapy. Addiction. 2008; 103:1660–1670. doi:10.1111/j.1360-0443.2008.02274.x. [PubMed: 18705691]

Liddle HA, Rowe CL, Dakof GA, Henderson CE, Greenbaum PE. Multidimensional family therapy for young adolescent substance abuse: Twelve-month outcomes of a randomized controlled trial. Journal of Consulting and Clinical Psychology. 2009; 77:12–25. doi:10.1037/a0014160. [PubMed: 19170450]

Liddle HA, Dakof GA, Henderson CE, Rowe CL. Implementation outcomes of Multidimensional Family Therapy-Detention to Community: A reintegration program for drug-using juvenile detainees. International Journal of Offender Therapy and Comparative Criminology. 2011; 55:587–604. doi:10.1177/0306624x10366960. [PubMed: 20427547]

Liddle HA, Dakof GA, Rowe CL, Henderson CE, Greenbaum PE, Wang W, Algerga A. Multidimensional Family Therapy as a community-based alternative to residential drug treatment: A randomized controlled trial. 2013 Manuscript submitted for publication.

Little, RJA.; Rubin, DB. Statistical analysis with missing data. 2nd ed.. Wiley; New York: 2002.

Lipsey, MW. Design sensitivity: Statistical power for experimental research. Sage; Newbury Park, CA: 1990.

Mak WWS, Law RW, Alvidrez J, Pérez-Stable EJ. Gender and ethnic diversity in NIMH-funded clinical trials: Review of a decade of published research. Administrative Policy in Mental Health. 2007; 34:497–503. doi:10.1007/s10488-007-0133-z.

Miranda J, Bernal G, Lau A, Kohn L, Hwang WC, LaFromboise T. State of the science on psychosocial interventions for ethnic minorities. Annual Review of Clinical Psychology. 2005; 1:113–142. doi:10.1146/annurev.clinpsy.1.102803.143822.

Muthén, L.; Muthén, B. Mplus user's guide. 7th ed.. Muthén & Muthén; Los Angeles, CA: 1998-2012.

National Institutes of Health. NIH policy and guidelines on the inclusion of women and minorities as subjects in clinical research. Amended. 2001

Rahdert, E. The Adolescent assessment/referral system manual. National Institute on Drug Abuse; Rockville, MD: 1991. DHHS Pub. No. (ADM) 91-1735

Rigter H, Henderson CE, Pelc I, Tossmann P, Phan O, Hendriks V, Schaub M, Rowe C. Multidimensional family therapy lowers the rate of cannabis dependence in adolescents: A randomised controlled trial in Western European outpatient settings. Drug and Alcohol Dependence. 2012; 130:85–93. doi:10.1016/j.drugdep.2012.10.013. [PubMed: 23140805]

Robinson WS. Ecological correlations and the behavior of individuals. American Sociological Review. 1950; 15:351–357.

Rose JS, Dierker LC, Hedeker D, Mermelstein R. An integrated data analysis approach to investigating measurement equivalence of DSM nicotine dependence symptoms. Drug and Alcohol Dependence. 2013; 129:25–32. doi:10.1016/j.drugalcdep.2012.09.005. [PubMed: 23021772]

Rowe C, Rigter H, Henderson C, Gantner A, Mos K, Nielsen P, Phan O. Implementation fidelity of Multidimensional Family Therapy in an international trial. Journal of Substance Abuse Treatment. 2013; 44:391–399. doi:10.1016/j.sat.2012.08.225. [PubMed: 23085040]

Sobell, LC.; Sobell, MB. Timeline follow-back: A technique for assessing self-reported alcohol consumption.. In: Litten, RZ.; Allen, JP., editors. Measuring alcohol consumption: Psychosocial

and biochemical methods. Humana Press; Totowa, NJ US: 1992. p. 41-72.doi: 10.1007/978-1-4612-0357-5-3

Stein DM, Deberard S, Homan K. Predicting success and failure in juvenile drug treatment court: A meta-analytic review. Journal of Substance Abuse Treatment. 2013; 44:159–168. doi:10.1016/j.ssat.2012.07.002. [PubMed: 22980448]

Strada MJ, Donohue B, Lefforge NL. Examination of ethnicity in controlled treatment outcome studies involving adolescent substance abusers: A comprehensive literature review. Psychology of Addictive Behaviors. 2006; 20:11–27. doi:10.1037/0893-164X.20.1.11. [PubMed: 16536661]

Sue S, Zane N, Hall GCN, Berger LK. The case for cultural competency in psychotherapeutic interventions. Annual Review of Psychology. 2009; 60:525–548. doi:10.1146/annurev.psych.60.110707.163651.

Szapocznik J, Prado G, Burlew AK, Williams RA, Santisteban DA. Drug abuse in African American and Hispanic adolescents: Culture, development, and behavior. Annual Review of Clinical Psychology. 2007; 3:77–105. doi:10.1146/annurev.clinpsy.3.022806.091408.

Wang, W.; Greenbaum, PE.; Henderson, C. Multiple calibrations in integrative data analysis: When needed, how many, and how to combine?. Department of Epidemiology and Biostatistics, University of South Florida; Tampa, FL.: 2013. Unpublished manuscript

Winters, KC.; Henly, GA. Personal experience inventory and manual. Western Psychological Services; Los Angeles: 1989.

Winters KC, Latimer WW, Stinchfield RD, Egan E. Measuring adolescent drug abuse and psychosocial factors in four ethnic groups of drug-abusing boys. Experimental and Clinical Psychopharmacology. 2004; 12:227–236. doi:10.1037/1064-1297.12.4.22. [PubMed: 15571439]

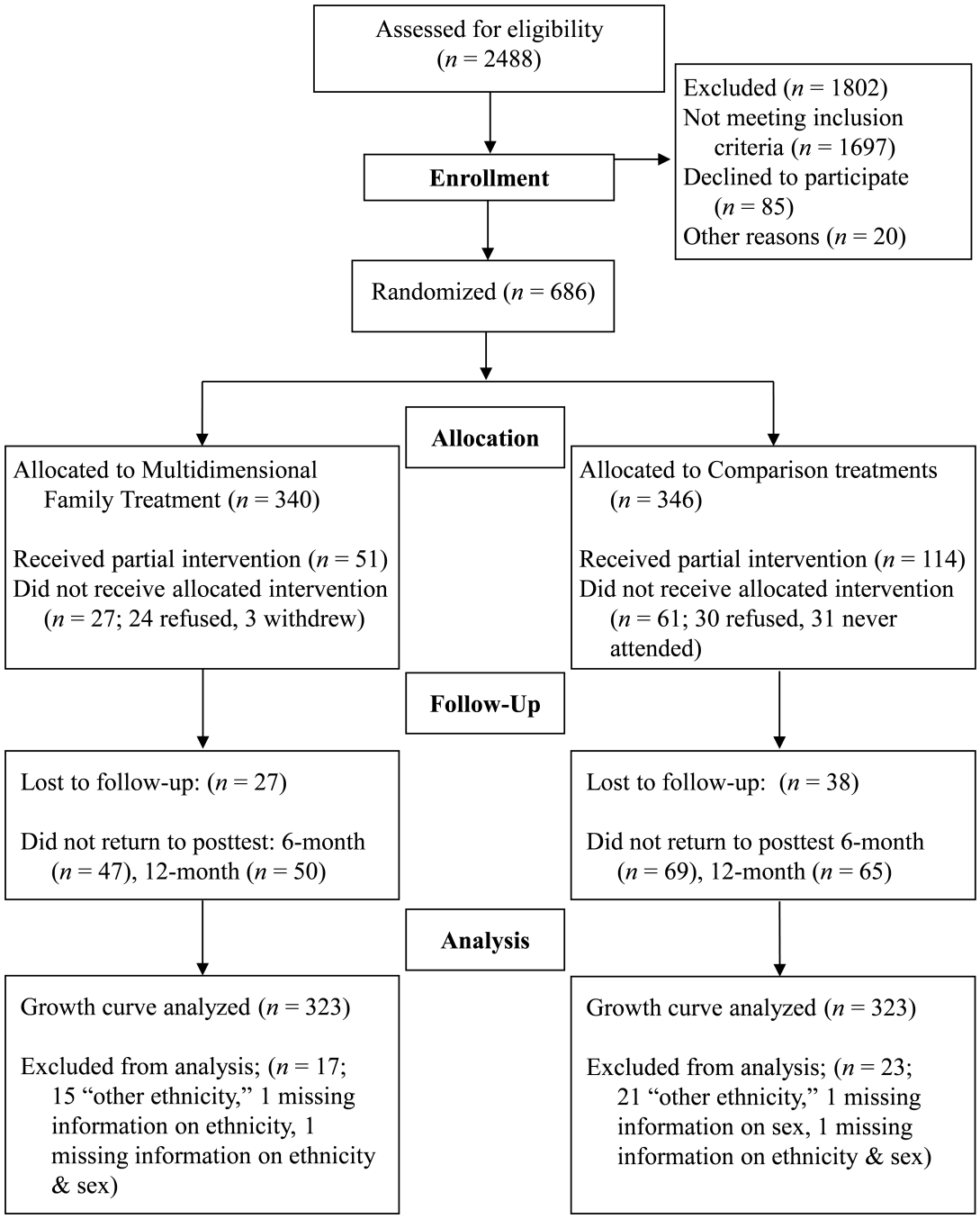

Figure 1.
Sampling and flow of participants through the five Multidimensional Family Therapy randomized controlled trials. Participant flow chart following Consolidated Standards of Reporting Trials (CONSORT) guidelines.

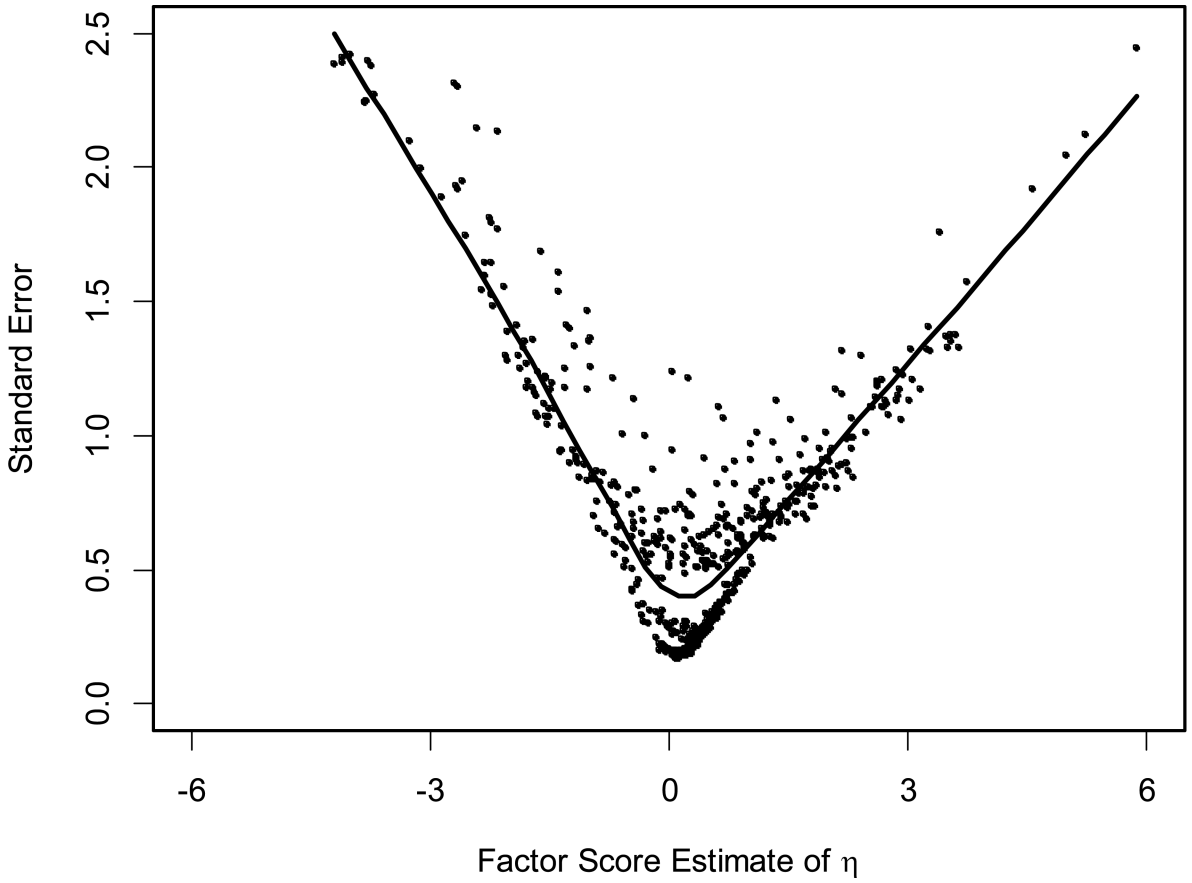

Figure 2.
Standard errors for maximum a posteriori (MAP) factor score estimates from calibration sample 17 as a function of the estimated factor score value.

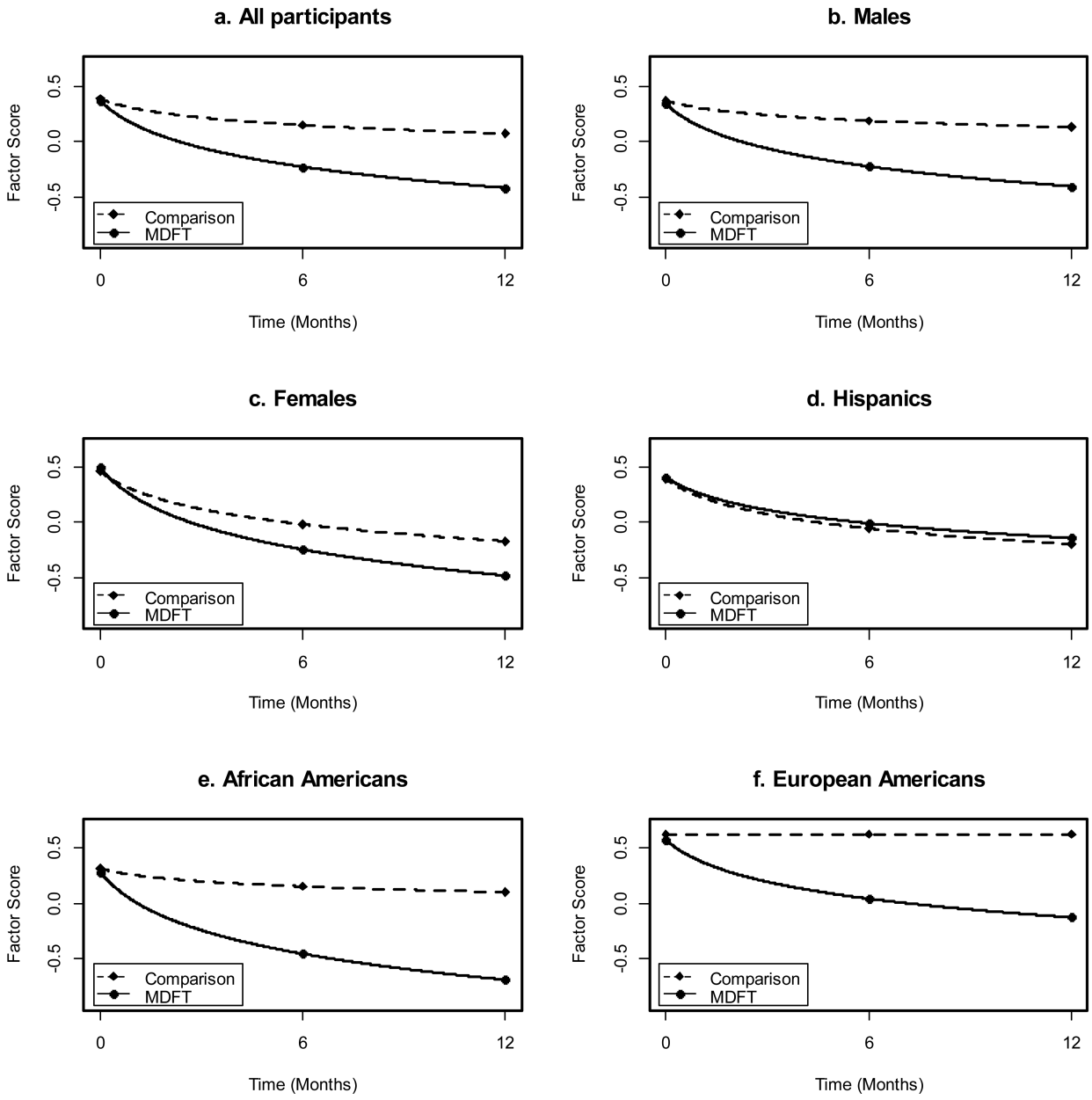

Figure 3.

Estimated treatment effects for (a) all participants (n = 646), (b) males (n = 522), (c) females (n = 124), (d) Hispanics (n = 225), (e) African Americans (n = 329), and (f) European Americans (n = 92) for the combined calibration samples. MDFT = Multidimensional Family Therapy.

Table 1

Summary of Sample Characteristics and Measures from the MDFT Randomized Controlled Trials

Study	Gender	Age *M* (SD)	Ethnicity	Baseline Impairment	Drug Use Measures	Assessments	Comparison
Liddle et al. (2008) (*n* = 224)	81% Male	15.4 (1.2)	72% AA 10% Hispanic 18% EA	• *M* = 14.03 times used 30 days • 87% drug dependence • *M* = 2.14 DSM diagnoses • 61% in justice system	• TLFB (30 days) • PEI	• Intake • Discharge • 6 Months • 12 Months	Individual Cognitive Behavioral Therapy
Liddle et al. (2009) (*n* = 83)	74% Male	13.8 (1.1)	38% AA 42% Hispanic 3% EA	• *M* = 3.65 days used 30 days • 16% drug dependence • 22% in justice system	• TLFB (30 days) • POSIT • Urinalysis	• Intake • 6 Weeks • Discharge • 6 Months • 12 Months	Group Therapy
Liddle et al. (2013) (*n* = 113)	74% Male	15.4 (1.1)	20% AA 68% Hispanic 8% EA	• *M* = 30.51 times used 30 days • 100% drug dependence • *M* = 3.83 DSM diagnoses • 81% in justice system	• TLFB (30 days) • PEI • POSIT • Urinalysis	• Intake • 2 Months • 4 Months • 12 Months	Multimodal Residential
Liddle et al. (2011) (*n* = 154)	83% Male	15.4 (1.1)	60% AA 22% Hispanic 18% EA	• *M* = 50.02 days used 90 days • 38% drug dependence • M = 2.71 DSM diagnoses • 100% in justice system[a]	• TLFB (90 days) • PEI • POSIT • Urinalysis	• Intake • 3 Months • 6 Months • 9 Months	Enhanced Services as Usual
Dakof et al. (2015) (*n* = 112)	90% Male	16.1 (1.0)	29% AA 60% Hispanic 4% EA	• *M* = 64.34 days used 90 days • 37% drug dependence • *M* = 1.77 DSM diagnoses • 100% in justice system[b]	• TLFB (90 days) • PEI • POSIT • Urinalysis	• Intake • 6 Months • 12 Months	Group Therapy

Note. AA = African American, EA = European American, DSM = Diagnostic Statistical Manual, TLFB = Timeline Follow-Back, PEI = Personal Experience Inventory, POSIT = Problem Oriented Screening Instrument for Teenagers

[a] Participants in detention at intake by study design.

[b] Participants in post-adjudication drug court at intake by study design