

Clinician Integrity in Multiple Family Groups: Psychometric Properties and Relationship with Schizophrenia Client and Caregiver Outcomes

Michael G. McDonell · Martina L. Rodgers ·
Robert A. Short · Diane Norell · Lori Pinter ·
Dennis G. Dyck

Published online: 12 May 2007
© Springer Science+Business Media, LLC 2007

Abstract Multiple family group treatment (MFGT) is an evidence-based cognitive behavioral psycho-education and problem-solving intervention for the management of schizophrenia. Despite empirical support for the efficacy of MFGT, there is no psychometric tool for measuring treatment integrity in MFGT. In addition, research has not examined the impact of treatment integrity on client and family outcomes. The psychometric properties of a clinician integrity measure and its relationship to client and family outcomes was investigated among 7 clinician dyads who provided treatment to 53 persons with schizophrenia and their families over 2 years. The treatment integrity measure was found to have adequate psychometric properties and was able to differentiate clinicians with low, moderate, and high treatment integrity. To investigate the association of integrity with outcomes, dyads were split into low/moderate and high MFGT integrity groups. It was hypothesized that participants who received high integrity MFGT would experience a greater reduction in client psychiatric symptoms and more positive psychological health in family caregivers, relative to persons who received low/moderate integrity MFGT. Accounting for pre-intervention levels of psychiatric symptoms and session attendance, clients who were provided high integrity MFGT demonstrated significantly lower post-randomization scores on two measures of psychiatric symptoms, relative to those who received low/moderate integrity MFGT. There was no effect of treatment integrity on family caregiver outcomes. Discussion focuses on the issues associated with measuring and assuring treatment integrity in clinical research and practice.

M. G. McDonell (✉)
Department of Psychiatry and Behavioral Medicine, University of Washington Medical School,
Odessa Brown Children's Clinic, Children's Hospital and Regional Medical Center, 2101 East
Yessler, Suite 100, Seattle, WA 98122-5999, USA
e-mail: michael.mcdonell@seattlechildrens.org

M. L. Rodgers · R. A. Short · D. Norell · D. G. Dyck
Washington State University, Spokane, WA 99210-1495, USA

L. Pinter
Spokane Mental Health, Spokane, WA 99202, USA

Keywords Family therapy · Psycho-education · Schizophrenia · Clinician integrity · Adherence · Competence · Psychotherapy outcomes

Multiple family group treatment (MFGT) is an adjunctive family psycho-education and support intervention for adults with schizophrenia. This approach and single family psycho-education (SFP) models have received extensive empirical support and are considered a best practice treatment for adults with schizophrenia (Dixon et al., 2001; Lehman & Steinwachs, 1998). Relative to SFP approaches, MFGT compares favorably in reducing psychiatric relapse and increasing functional outcomes (McFarlane, Link, Dushay, Marchal, & Crilly, 1995; McFarlane, Lukens, et al., 1995). When compared to a standard care group, MFGT has shown reductions in negative symptoms, inpatient service utilization, and caregiver distress in an outpatient community mental health setting (Dyck et al., 2000; Dyck, Hendryx, Short, Voss, & McFarlane, 2002; Hazel et al., 2004; McDonnell, Short, Hazel, Berry & Dyck, 2006). As a result of this empirical support of family psycho-education in general, and MFGT the format specifically, MFGT is currently being disseminated throughout the United States through the Substance Abuse and Mental Health Administration's (SAMHSA; 2003) Toolkit for Empirically Supported Treatments. This Toolkit provides information for the successful implementation of MFGT and other empirically supported pharmacological and psychosocial treatments.

MFGT involves four phases: joining with the client and family, a psycho-educational workshop, relapse prevention, and vocational and social skills enhancement (McFarlane, 2002) and is delivered by two clinicians to groups of 5–8 families for 2–3 years. Treatment goals are to decrease client symptoms, psychiatric relapse, and improve social and vocational skills. Treatment strategies include psycho-education and problem-solving in order to teach family members and clients to reframe, reinterpret, and re-attribute client behaviors in a manner that results in reduced expressed emotion and improved family communication skills.

While MFGT is based upon earlier single-family psycho-education and support interventions for schizophrenia, many of the components of MFGT are similar to cognitive behavioral therapy (CBT) for schizophrenia. Both approaches emphasize the importance of establishing a strong therapeutic alliance with clients as a basis for psycho-education and providing new strategies to aid in symptom reduction, illness relapse, and recovery from mental illness (Garety, Fowler, & Kuipers, 2000; Kingdon & Turkington, 2005). The family focus of MFGT and the reliance on behavioral and problem-solving strategies differentiate MFGT from CBT, some of the originators of CBT have advocated that family interventions are an important accompaniment to individual cognitive therapy (Turkington, Dudley, Warman, & Beck, 2004).

Treatment integrity

Clinician integrity refers to the extent to which a treatment is implemented as designed. Waltz, Addis, Koerner, and Jacobson (1993) described clinician integrity as consisting of adherence and competence. Adherence refers to the degree that clinicians engage in prescribed behaviors and avoid proscribed behaviors of an intervention and is typically assessed using checklists of prescribed and proscribed behaviors (e.g., Beck, Rush, Shaw, & Emory, 1979). Clinician competence is defined as the art or skill used in the implementation of prescribed behaviors (e.g., timing a technique effectively).

Hypothetically, increased integrity should lead to increased efficacy and effectiveness. Waltz et al. also assert that contextual factors, particularly the difficulty of the client receiving treatment (e.g., clients with severe symptoms), can influence treatment integrity.

Treatment integrity has been linked to positive therapeutic outcomes in cognitive behavioral, psychodynamic/interpersonal, and multi-systemic therapies (Barber, Crits-Christoph, & Luborsky, 1996; DeRubeis & Feeley, 1990; Feeley, DeRubeis, & Gelfand, 1999; Huey, Henggeler, Brondino, & Pickrel, 2000; Luborsky, McLellan, Woody, O'Brien, & Auerbach, 1985; O'Malley et al., 1988; Rounsaville et al., 1987). While evidence supporting this relationship is not uniformly positive (e.g., Huppert et al., 2001; Markowitz, Spielman, Scarvalone, & Perry, 2000), overall existing literature supports the relationship between treatment integrity and positive psychotherapy outcomes. However, limited research has assessed the relationship between treatment integrity and client outcomes in family psychotherapies, such as MFGT (Huey, Henggeler, Brondino, & Pickrel, 2000; Weisman et al., 1998), and no previous study has assessed the relationship between treatment integrity and family outcomes, such as family psychological functioning.

Treatment integrity and MFGT

Currently, the SAMHSA Toolkit incorporates a measure of MFGT integrity developed by McFarlane and colleagues, the originators of the intervention. This measure assesses structural components of the intervention relevant to dissemination (e.g., the number of sessions that must be provided, correct sequencing of sessions), rather than in-session therapist behaviors. Previous research has typically measured treatment integrity by assessing adherence and competence ratings of in-session therapist behaviors (e.g., Feeley et al., 1999; O'Malley et al., 1988). Therefore, while the MFGT integrity measure developed for the SAMHSA toolkit may be a useful tool, there was no empirically validated tool to measure therapist competence and adherence in MFGT.

The first part of this paper addresses the psychometric properties of the *Multiple Family Group-Adherence and Competence Checklist* (MFG-ACC), which is a measure of in-session treatment integrity (adherence and competence) for the problem-solving phases of MFGT. The first part of the paper addresses the psychometric properties of this tool among seven clinician dyads providing MFGT as part of a clinical effectiveness study. In the second part, the relationship between ratings of treatment integrity and client and caregiver outcomes was investigated among 53 client/caregiver dyads who received MFGT. We hypothesized that the MFG-ACC would demonstrate adequate psychometric properties and that clients and caregivers who received high integrity MFGT would experience positive outcomes, relative to those who did not receive high integrity MFGT. Research evidence for these hypotheses support the use of the MFG-ACC as a reliable and valid measure of in-session MFGT integrity in research and clinical domains.

Method

Participants

The sample consisted of 53 persons with schizophrenia and related psychotic disorders and their caregivers who were randomized to the standard care plus MFGT arm of a

clinical trial ($n = 106$) designed to assess the effectiveness of this intervention in a community mental health setting. All clients were required to have 5 h of contact per week with one specified family caregiver. All participants provided informed consent and institutional review board approval was provided by all appropriate agencies. Each client identified one caregiver who consented to participate in this investigation. Clients were aged 19–58 years with *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-IV), diagnoses of schizophrenia ($n = 36$), schizoaffective ($n = 16$), or other psychotic disorders ($n = 1$; American Psychiatric Association, 1994), as measured with the *Structured Clinical Interview for the DSM-IV, Psychotic Disorders Version* (SCID; First, Spitzer, Gibbon, & Williams, 1995). Caregivers were typically European-American mothers. Table 1 provides a summary of client and caregiver demographic and clinical characteristics at study entry.

Study protocol

This examination of treatment integrity was based on a review of videotaped treatment sessions by MFGT trainees and supervisors. Participants were enrolled between 1995 and 2000 in seven cohorts. Randomization to MFGT plus treatment as usual versus treatment as usual was conducted within seven cohorts. Participants received treatment

Table 1 Descriptive analyses of participants at randomization ($n = 53$)

Variable	Client		Caregiver	
	Percent	Mean (SD)	Percent	Mean (SD)
Age		32.8 (9.4)		51.3 (12.1)
Gender				
Male	77.9		16.5	
Female	22.1		83.5	
Race				
White	92.6		91.8	
Non-white	7.4		8.2	
Diagnosis				
Schizophrenia	68			
Paranoid	45			
Undifferentiated	17			
Disorganized	6			
Schizoaffective	30			
Other psychotic disorder	2			
BPRS total score		40.2 (10.6)		
Negative symptoms (mSANS)		8.7 (3.0)		
Psychiatric hospitalization 1 year pre-randomization	70.0			
Illness duration (Years)		9.0 (8.0)		
History of substance abuse	74.0			
Relationship to client				
Parent			73.2	
Sibling			12.4	
Spouse/Partner			7.2	
Child			3.1	
Other relative			2.1	
Other non-relative/friend			2.1	

Note. Brief psychiatric rating scale (BPRS), modified scale for the assessment of negative symptoms (mSANS)

as usual and MFGT through the local outpatient community mental health center. Treatment as usual consisted of medication management, case management, group psychotherapy, treatment for substance abuse as well as clubhouse and other social and vocational services. The MFGT model attempted to faithfully replicate the intervention developed by McFarlane et al. (1991). Each problem-solving session was approximately 90 min and consisted of an initial 10-min socialization period, a 45-min go around, followed by problem-identification and problem-solving, and concluded with a 5-min socialization period. See McFarlane (2002) for a detailed description of this intervention. MFGT clinicians who were provided with training and supervision were asked to videotape all sessions. Videotapes were used for supervision purposes and then were rated after completion of the study to assess clinician integrity.

Clinician dyads and clinical supervision

All clinicians had extensive experience providing mental health services to adults with serious and persistent mental illness. However, they varied in formal education level (three bachelor, seven master, and one doctoral level clinician) and professional background (psychology, social work, and psychiatric nursing). Three of the clinicians provided treatment for two cohorts, while the other eight provided treatment for only one cohort. Clinicians were required to complete extensive reading and participate in a 2-day intensive training workshop, which demonstrated the structure and process of MFGT and provided role-play experience. After training, all clinicians received monthly supervision by the originators of the intervention and locally trained MFGT supervisors, utilizing videotape. Both locally trained supervisors were also MFGT clinicians in the study. To protect the confidentiality of clinicians, the seven dyads were randomly labeled with consecutive letters (i.e., dyads A, B, C, D, E, F, and G).

Measures and procedure

Videotape selection

The study protocol required that all MFGT sessions be videotaped for supervision and quality assurance. In order to provide an accurate assessment of clinician integrity, a total of six videotapes from each dyad selected to be rated from 3, 6, 9, 12, 15, and 21 months post-randomization. All clinician dyads, except for dyad A, had over 30 available videotapes. Potential videotaped sessions were reviewed for audio and visual quality by two persons who did not conduct MFG-ACC ratings. Only videotapes that had adequate audio and video quality were rated. Videotapes within 2 months of the above time periods were obtained for all dyads, except for dyad A. Dyad A had 6 videotapes available corresponding to 9, 12, 18, 19, 21, and 23 months post-randomization. Dyad A had only six videotaped sessions available, as they did not adhere to study videotape procedures (i.e., they did not videotape all sessions).

MFGT adherence and competence checklist

The MFG-ACC (Appendix A) was developed by the fourth author, in consultation with the last author. The content was based on the MFGT manual developed by McFarlane et al. (1991) and the format was modeled after the *Competency Checklist for Cognitive*

Therapists (Beck et al., 1979). As problem-solving is central to McFarlane's model, the MFG-ACC targeted the integrity of clinician behaviors during the group problem-solving sessions of MFGT. It is divided into subscales corresponding to the (1) socialization, (2) go around, (3) problem identification, and (4) problem-solving phases of the group sessions. Clinician adherence and competence ratings were assigned for each subscale.

An initial pool of adherence items was developed. Each adherence item described a prescribed clinician behavior. Clinicians received one point if they engaged in a prescribed behavior or appropriately omitted the behavior and no points if they engaged in a proscribed behavior. Clinician adherence scores were summed within each subscale to obtain adherence subscale scores and across subscales to obtain overall adherence scores. The final version of the MFG-ACC included 43 adherence items, resulting in a total score ranging from 0 to 43.

Competence was also rated for each phase of the group sessions by using two, 7-point Likert scaled questions (Appendix A). The first competence item of each subscale read: "how skillfully did the clinicians conduct the [particular problem-solving phase, e.g. socialization]?" The second question assessed clinicians' ability to work together and read: "how skillfully did the clinicians work together during [particular phase, e.g., socialization]?" The second competence question was only rated when both clinicians were present. Anchors for both questions ranged from one ("unskillfully") to seven ("skillfully"). Anchors were defined generally because further detail required information that overlapped with adherence items. The two competence items were averaged within each subscale, and then averaged across subscales to produce a total competence score, which ranged from one to seven points.

The first author and the two onsite clinical supervisors of the project completed MFGT adherence and competence ratings. All raters had a master's degree in a mental health field. The two study clinical supervisors, raters 1 and 2, had over 20 years of experience providing treatment for persons with SPMI, as well as 8 years of experience implementing and supervising MFGT. At the time of these ratings, the first author (rater 3) was a doctoral student in clinical psychology with 4 years of experience providing mental health treatment. Raters 1 and 2 did not rate groups in which they provided treatment. Rater 1 rated 23 videotaped sessions, rater 2 rated 20 videotaped sessions, and rater 3 rated 29 MFGT sessions. Thirty videotapes were rated by two raters in order to assess inter-rater agreement and reliability, resulting in a total of 72 ratings. Videotapes were randomly ordered before being rated. To calibrate ratings, raters first rated three pilot sessions. They independently rated subscales and then discussed their adherence and competence ratings. To prevent rater drift, the raters rated half of their assigned videotapes and then met for discussion. At that time, they rated an additional videotape to assure reliability.

Client and family outcome measures

Psychiatric symptoms were assessed as client outcomes. Measures of psychiatric symptoms included the sum of items from the *Brief Psychiatric Rating Scale* (BPRS; Overall & Gorham, 1962) and the sum of subscale scores from the *modified Scale for the Assessment of Negative Symptoms* mean of subscale scores (mSANS; Andreasen, 1983). Twenty-three items from the BPRS were summed to assess positive and negative symptoms, as well as other symptoms commonly experienced by persons with severe psychiatric illness (possible range = 23–161). Although the mSANS contains 30 items,

only items assessing global status were summed to create the index. These items assessed alolia, affective blunting, asociality, anhedonia, avolition and inattention (possible range = 5–25). For each measure of psychiatric symptoms, monthly administrations were averaged across two 4-month time periods (10–14 months post-randomization and 22–26 months post-randomization) to correspond with the timeline of caregiver assessment. The mSANS and BPRS were administered by trained raters blind to treatment condition and treatment integrity ratings.

Caregiver outcomes were assessed at randomization, and then again at 1 and 2 years post randomization. Family caregiver psychological distress and resource variables were empirically derived from analyses conducted by Hazel et al. (2004), which demonstrated high correlations between measures comprising these composites. Caregiver distress consisted of the standardized sum of scores from the *Perceived Stress Scale* (Cohen, Kamarck, & Mermelstein, 1983), the Anger-In subscale of the *Anger Expression Scale* (Spielberger et al., 1985), the Global Distress Index of the *Center for Epidemiological Studies Depression Scale* (Radloff, 1977), and the State subscale of the *State-Trait Anxiety Inventory* (Spielberger, Gorsuch, & Lushene, 1970). The family caregiver resources composite variable was calculated by standardizing and summing the overall *Social Support Questionnaire* (Sarason, Levine, Basham, & Sarason, 1983) and the *Interpersonal Support Evaluation List* (Cohen, Mermelstein, Kamarch, & Hoberman, 1985) scores, and then subtracting the standardized Passive Coping subscale score from the *Revised Ways of Coping Checklist* (Vitaliano, Russo, Carr, Maiuro, & Becker, 1985). Please see Hazel and colleagues (2004) for a more complete theoretical and empirical description of these composites.

Client and family attendance

Client and family attendance was estimated through use of the six videotapes of each clinician dyad that were chosen for ratings. Each client and family member was assigned one point for every session they attended in order to reflect the overall attendance of the dyad. Therefore, each client/family member dyad could receive 0–12 points on this measure. The mean number of sessions attended was 5.1 (SD = 4.1). A significant reduction in overall attendance across treatment sessions was observed, $F(5,50) = 2.28$, $P < .05$. When investigated across session, persons receiving low/moderate and high integrity MFGT did not differ in attendance. Attendance was not significantly correlated with client and family difficulty.

Data analytic plan and power analyses

Measuring the psychometric properties of the MFG-ACC

Inter-rater agreement on MFG-ACC adherence items was assessed using a modified κ coefficient (Fleiss, 1971), which allowed for assessment of the overall agreement of 3 raters where 2 raters rated each of 30 sessions. Inter-rater reliability between the 3 rating dyads (1 & 2, 1 & 3, 2 & 3), was assessed using intraclass correlations (ICC; Shrout & Fleiss, 1979), which were calculated for each competence item, subscale, and the overall score using the 30 videotaped sessions. Internal consistency of the MFG-ACC adherence subscale and overall scores were investigated using the Kuder–Richardson Formula 20 across all 72 measurements. Internal consistency of competence

items was calculated using Cronbach's alpha across 42 observations, including the average item ratings for the 30 sessions that were rated twice.

Impact on client and family caregiver outcomes

The effect of treatment integrity on outcomes was investigated by comparing those participants who received low/moderate integrity ($n = 23$) to those who received high integrity MFGT ($n = 30$). Clinician dyads were separated into groups based on their performance on the MFG-ACC (see below). Low and moderate integrity dyad were combined into one group for these analyses in order to form relatively equal comparison groups. In order to investigate outcomes, all clients and caregivers were assigned to groups based on their clinician dyad's performance on the MFG-ACC. Repeated measures analysis of covariance (ANCOVA) was used to investigate the main effect of MFGT integrity (high versus low/moderate integrity MFGT groups) and the interaction of MFGT integrity and time (years 1 & 2 post-randomization) on each of the four outcome variables across the 2-year intervention period, accounting for baseline levels of outcome variables and attendance. Power analyses revealed that, assuming a 2 (high versus low/moderate integrity MFGT) by 2 (number of post-randomization time periods) repeated measures design, an $\alpha < .05$, and a correlation of $r = .30$ within the dependent measure across time, there was an 80% chance of detecting a difference between group means of 0.29 standard deviation units and a difference of 0.27 standard deviation units for the interaction between group and time for 53 MFGT client/family member dyads.

Results

Measuring the psychometric properties of the MFG-ACC

Psychometric properties

Interrater reliability Across the 43 adherence items, inter-rater agreement ranged from $\kappa = -.11$ to $\kappa = 1.00$ (Mean $\kappa = .50$). Raters attained perfect agreement on three items and agreement below chance on eight items. Average subscale inter-rater agreement is shown in Table 2. Inter-rater reliability of the eight competence items ranged from below chance to ICC = .94. Intraclass correlations for each rating dyad are presented in Table 3. Raters one and three had relatively low reliability across competence domains, while raters one and two were most consistent (Table 3).

Internal consistency and scale structure Each MFG-ACC item was examined. Items that did not demonstrate adequate variance and/or had very low inter-rater reliability were removed from the scale. This allowed for a measure that maximized reliability and individual differences between clinician dyads. MFG-ACC competence scores demonstrated adequate variance and adequate inter-rater reliability. Therefore, all item, subscale, and overall competence scores were included, without modification, in the final version of the MFG-ACC. The overall competence scale had high internal consistency, $\alpha = .94$.

Adherence ratings had limited variance. Therefore, adherence items were only included if they had inter-rater agreement above chance and at least 10% of persons rated in prescribed, appropriately omitted, or proscribed categories. This resulted in an

Table 2 Item inter-rater agreement for MFG-ACC adherence subscales

Subscale	Number of items	Lowest κ	Highest κ	Mean κ
Socialization	9	-.08	1.00	0.53
Go around	19	-.07	.89	0.35
Problem identification	5	-.01	1.00	0.62
Problem-solving	10	.34	1.00	0.68

Table 3 Intra-class correlations of MFG-ACC competence items, subscales, and overall scores

	Raters		
	1 & 2 ($n = 7$)	1 & 3 ($n = 11$)	2 & 3 ($n = 12$)
Socialization	.81	.77	.59
Question 1	.61	.54	.60
Question 2	.94	.45	.72
Go Around	.82	.32	.61
Question 1	.87	.45	.35
Question 2	.53	.00	.79
Problem identification	.83	-.17	.51
Question 1	.85	-.06	.48
Question 2	.51	-.22	.62
Problem-solving	.90	.10	.75
Question 1	.90	-.11	.61
Question 2	.90	.05	.79
Total Competence	.89	.48	.74

adherence scale with 24 of the original 43 items. The revised adherence scale had an internal consistency of $\alpha = .87$. Five of the nine items met criteria for inclusion on the socialization subscale, $\alpha = .57$. Nine of a possible 19 items met criteria for inclusion on the go around subscale, $\alpha = .74$. The modified problem identification subscale contained three of the five possible items, $\alpha = .73$. Because the problem identification subscale had only 3 items and demonstrated both a theoretical and empirical, $r(42) = .73$, $P < .01$, link to the problem-solving subscale, these items were combined with 7 of the 10 problem-solving subscale items, $\alpha = .86$.

Modified adherence subscale correlations ranged in magnitude from $r = .47$ to $.60$ with an average magnitude of $r = .53$. All subscale correlations were significant, $P < .01$. In addition, total adherence and competence total, $r(42) = .90$, $P < .01$, and subscale scores ($r = .83$ to $.90$) were significantly correlated. Because of the magnitude of these correlations, adherence and competence scores were standardized and summed to produce an overall MFGT integrity score.

MFGT clinician performance

Overall, MFGT clinician performance demonstrated a high level of treatment integrity. Across the 42 MFGT sessions, clinicians had an average overall adherence score of 35.34 ($SD = 7.66$) representing prescribed behaviors on 82% of MFG-ACC items. Clinicians attained an average overall competence score of 5.14 ($SD = 1.31$), placing them in the “skillful” half of the seven-point scale. All competence items were negatively skewed.

All dyads, except for dyad A, had average adherence scores above 36 points. Four dyads attained adherence scores above 88%, 2 dyads attained scores of 81–84%, and dyad A attained 49% adherence. Dyad A clinicians were the only MFGT clinicians who did not attain an adherence rating above the overall mean rating of 35.34. A significant effect for dyad was observed using a one way analysis of variance, $F(6,35) = 9.56$, $P < .01$. Post-hoc analyses indicated that only comparisons of dyad A clinicians to other MFGT clinicians were significant.

Average overall competence scores demonstrated a similar pattern. All dyads, except for dyad A, attained average competence scores of 5.15–6.75. As with their average adherence scores, dyad A clinicians averaged a low overall competence score of 2.66 ($SD = .74$). This difference resulted in a significant dyad effect, $F(6,35) = 14.77$, $P < .01$. Post-hoc comparisons demonstrated that dyad A had significantly lower competence scores relative to other clinicians.

Across dyads, adherence and competence subscale scores followed a pattern similar to overall scores. Clinician adherence and competence across sessions was investigated using one-way ANOVA's. No significant pattern of change in overall adherence and competence scores was observed across sessions.

Impact on client and family caregiver outcomes

Client psychiatric symptoms

No significant group differences across client outcomes were observed at baseline. When post-randomization overall BPRS scores were investigated using repeated measures ANCOVA, the main effects of MFGT integrity, $F(1,34) = 8.2$, $P < .01$, and client difficulty (baseline BPRS total score), $F(1,34) = 66.6$, $P < .01$, were both statistically significant (Fig. 1). Across the 2-year intervention time period, clients who received high integrity MFGT (Mean = 35.4, SEM = 1.1) averaged significantly lower scores on the BPRS than those receiving low/moderate integrity MFGT (Mean = 40.4, SEM = 1.3; Fig. 1).

Similarly, when post-randomization mSANS scores were analyzed, significant main effects of MFGT integrity $F(1,34) = 4.4$, $P < .05$, (Fig. 2) and client difficulty (baseline

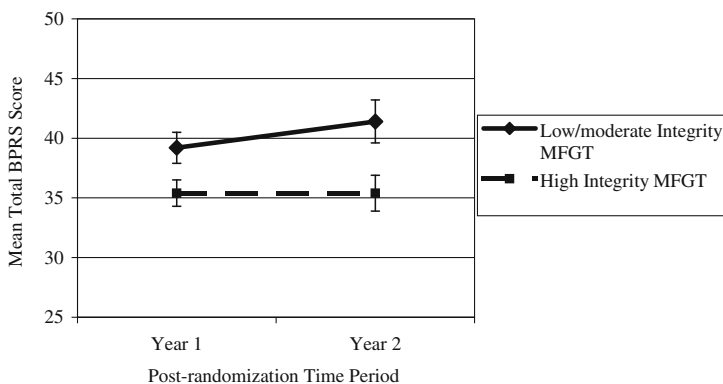


Fig. 1 Differences in mean total brief psychotic rating scale (BPRS) scores between low/moderate and high MFGT integrity groups, $F(1,34) = 8.2$, $P < .01$

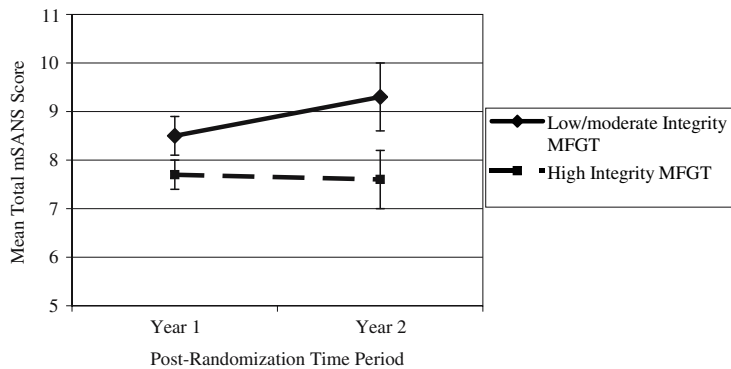


Fig. 2 Differences in mean total modified scale for the assessment of negative symptoms (mSANS) scores between low/moderate and high MFGT integrity groups, $F(1,34) = 4.4$, $P < .05$

mSANS scores), $F(1,34) = 29.8$, $P < .01$, were observed. Clients who received high integrity MFGT had significantly lower post-randomization negative symptoms (Mean = 7.7, SEM = .4), relative to those who received low/moderate quality MFGT (Mean = 8.9, SEM = .4; Fig. 2).

Family caregiver outcomes

Results of repeated measures ANCOVAs demonstrated no statistically significant differences between caregivers who received low/moderate and high levels of MFGT integrity. Only caregiver difficulty (i.e., baseline measurements of the respective outcome variable) was a significant predictor of increased caregiver distress, $F(1,33) = 58.4$, $P < .01$, and reduced resources, $F(1,27) = 37.4$, $P < .01$.

Discussion

Psychometric properties of the MFG-ACC

Inter-rater reliability and internal consistency

Inter-rater agreement of adherence items was acceptable (mean $\kappa = .50$). The agreement of raters across most adherence items was consistent with literature suggesting that less experienced raters can accurately assess adherence (Waltz et al., 1993). In terms of competence ratings, the differential inter-rater reliability across rating dyads supports the suggestion of Waltz et al., that only clinicians with significant experience in the intervention of study can accurately rate competence. For instance, raters one and two, experienced MFGT clinicians and supervisors, had the highest inter-rater reliability, while ICCs between rater three, a less experienced MFGT clinician, and the other raters were lower. Improvements, such as clearly defining competence questions and utilizing experienced raters, may improve the inter-rater reliability of competence items.

The internal consistency of the overall adherence and competence scales were high, $\alpha = .87$, $\alpha = .94$, respectively, demonstrating that these scores reflect unitary constructs.

After modifications to the MFG-ACC were made, adherence subscales, with the exception of socialization, demonstrated internal consistency above $\alpha = .73$. The low internal consistency of the socialization subscale, $\alpha = .57$, likely reflects the fact that this subscale consisted of five items and two of these items were related to the duration of socialization, not specific socialization behaviors. In addition, the three adherence subscales were moderately correlated, which verifies the a priori theoretical structure of the MFG-ACC and supports the combination of the problem identification and problem-solving subscales.

The magnitude of the correlation between adherence and competence, $r = .90$, indicated that they assess a unitary measure of integrity, rather than related constructs. The high correlation of adherence and competence is consistent with previous literature (Blackburn et al., 2001; Dobson, Shaw, & Vallis, 1985; Feeley et al., 1999; Haddock et al., 2001; Hill, O'Grady, & Elkin, 1992; Schoenwald, Henggeler, Brondino, & Rowland, 2000; Shapiro & Startup, 1992; Svartberg, 1989; Waltz et al., 1993). While Waltz et al. suggested the dichotomy of adherence and competence over 13 years ago, subsequent investigations, including this study, have failed to find evidence for this distinction (Blackburn et al.; Feeley et al.; Haddock et al.; Schoenwald et al.). Nonetheless, it is possible that non-specific clinician behaviors and the therapeutic alliance are important aspects of competence. Unfortunately, these elements were not explicitly measured as part of the assessment of clinician integrity.

Clinician performance

Overall, clinicians in this 2-year clinical trial were rated as 82% adherent to the MFGT model, with an average competence score of 5 on the 7-point scale. Comparisons of clinician adherence and competence in this investigation to previous treatment integrity studies is difficult, as previous investigations utilized disparate measures of treatment integrity (e.g., Barber & Crits-Christoph, 1996; Weisman et al., 1998). Although three groups of MFGT clinicians were differentiated using the MFG-ACC, only dyad A demonstrated adherence and competence scores significantly below other clinicians. Except for Dyad A, results indicate a consistent level of treatment integrity and suggest a successful supervision protocol. While only Dyad A was found to be non-adherent, the MFG-ACC was able to differentiate three levels of therapist performance. Therefore, even in this relatively small sample of clinicians, the MFG-ACC was able to differentiate dyads with varying levels of adherence and competence.

Typically, investigations of clinician integrity have been limited to individual psychotherapy (e.g., 10–20 sessions; Hill et al., 1992; Hoglend & Piper, 1995; Svartberg, 1989), with the longest assessment of integrity being 9 months (Weisman et al., 1998). The 2-year duration of this intervention allowed for an assessment of the stability of adherence and competence across time. The finding that the clinicians' MFG-ACC scores were consistent across time demonstrates that clinicians maintained their respective levels of adherence and competence but did not improve across time, as might be expected (e.g., practice & supervisor effects; Henry, Schacht, Strupp, Butler, & Binder, 1993; James, Blackburn, Milne, & Reichfelt, 2001; Multon, Kivlighan, & Gold, 1996). However, improvement may have been difficult given the intensive training received prior to MFGT implementation, which may have resulted in moderate to high levels of treatment integrity from the start of the intervention.

Treatment integrity and outcomes

Not only did the MFG-ACC demonstrate adequate internal consistency, inter-rater reliability, and identification of dyads demonstrating varying levels of adherence and competence, it was also found to be associated with differential client outcomes. The association of MFG-ACC with differential client symptoms supports the validity of the measure.

On a practical level, the observed relationship between treatment integrity and client outcomes should be of particular interest to administrators and clinicians wishing to implement MFGT in a clinical setting. As previously mentioned, MFGT is being widely disseminated throughout the United States, as part of the SAMSHA Evidence-based Mental Health Toolkit (2003). This study indicates that while there is a cost of time and money associated with the provision of supervision, training, and integrity monitoring, it is a worthwhile investment since MFGT integrity is necessary for improving client outcomes. Substantial training and supervision by mental health professionals who have experience delivering the intervention in an adherent and competent fashion may increase the likelihood that MFGT is delivered as prescribed. Tools such as the MFG-ACC, as well as other measures of treatment integrity, are helpful in improving the quality of feedback and supervision.

To our knowledge, this is the first investigation of the relationship between treatment integrity and family caregiver outcomes. While MFGT integrity appeared to be related to client psychiatric symptoms, no relationship between integrity and family caregiver outcomes was observed. Although somewhat surprising, the lack of a relationship between MFGT integrity and family outcomes may be related to a number of factors. First, the explicit focus of MFGT is on the illness of the client, rather than on the psychological well being of the caregiver. MFGT is designed to improve client functioning through psycho-educational and problem-solving sessions (e.g., reduce expressed emotion through communication training). Thus, any impact of MFGT on family caregivers is designed to be a means to achieve improved functioning of the client. A review of MFG-ACC items demonstrates the explicit focus of problem-solving sessions on client functioning (Appendix A). Therefore, the link between family functioning and MFGT integrity, as measured by the MFG-ACC, may be less direct compared to the relationship between integrity and client outcomes.

In addition, this study did not measure family expressed emotion and communication skills as caregiver outcomes, even though research has consistently found these family variables to be related to client psychiatric symptoms (Kavanagh, 1992; Vaughn, Snyder, Jones, Freeman, & Falloon, 1984). As a result, these family constructs might have been more directly impacted by MFGT integrity. It is also acknowledged that the relatively small sample size of this study may have limited our ability to detect a small to moderate effect size. Based on the relatively small sample size of this investigation and the lack of research in this area, future research should investigate the impact of treatment integrity on caregiver outcomes that are explicitly addressed by the treatment (e.g., expressed emotion).

Study limitations

While this report summarizes the initial psychometric properties of the MFG-ACC and its association with participant outcomes, interpretation is limited by a number of

factors. Although clinician integrity was a quality control issue in this study, the clinical trial was not originally designed as a prospective investigation of treatment integrity. Investigations of treatment integrity among individual psychotherapy approaches typically include a larger number of clinicians than in the present study, allowing for the possibility of greater variance in treatment integrity and increased external validity. Another weakness of the current study is the atypical method of scale development. The MFG-ACC was initially created as a tool for clinical supervision. Therefore, traditional scientific scale construction was not possible. Future MFG-ACC investigations should increase the number of clinicians rated and modify the tool to maximize variance between clinicians.

Beyond the methodological limitations of this study, the generalizability of the MFG-ACC to other CBT approaches and clinical populations deserves further investigation. For instance, only small modifications are necessary in order to use the MFG-ACC with similar problem-solving family treatments for persons with schizophrenia (e.g., Anderson, Reiss, & Hogarty, 1986).

Conclusion

The results of the current investigation indicate that adherence and competence in delivering MFGT problem-solving sessions can be measured reliably using the MFG-ACC. In addition, MFG-ACC scores identified three groups of clinicians who attained differential levels of treatment integrity. Clinician performance on the MFG-ACC was also correlated with client outcomes in this family cognitive behavioral problem-solving and psycho-education model. Results demonstrated a negative correlation between treatment integrity and client outcomes. Specifically, clients who received high integrity MFGT demonstrated significantly lower ratings on two measures of psychiatric symptoms (BPRS & mSANS) across the study period, relative to those who received low/moderate integrity MFGT. The relationship was observed despite statistically accounting for the important contextual variables of attendance and client difficulty. These results provide predictive validity to the MFG-ACC. In addition, results are consistent with the majority of previous treatment integrity studies that also observed a positive correlation between higher treatment integrity and improved client outcomes (Barber et al., 1996; DeRubeis & Feeley, 1990; Feeley et al., 1999; Huey et al., 2000; Luborsky et al., 1985; O'Malley et al., 1988; Rounsaville et al., 1987).

Results of this investigation suggest that the MFG-ACC is a valuable tool for supervisors and clinicians who are implementing MFGT and other family psycho-education and problem-solving interventions. As MFGT is being widely disseminated throughout the United States (SAMHSA, 2003), the MFG-ACC could be an important tool for accurate and reliable dissemination of the clinical treatment model. To assure accurate and effective dissemination of the intervention, the MFG-ACC could be used with the MFGT fidelity measure developed by McFarlane and colleagues (i.e., the extent to which systemic components of the intervention are implemented as prescribed) for the SAMSHA Toolkit.

References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Anderson, C. M., Reiss, D. J., & Hogarty, G. E. (1986). *Schizophrenia and the family: A practitioner's guide to psychoeducation and management*. New York: The Guilford Press.
- Andreasen, N. C. (1983). *The scale for the assessment of negative symptoms*. Iowa City, Iowa: The University of Iowa.
- Barber, J. B., & Crits-Christoph, P. (1996). Development of a therapist adherence/competence rating scale for supportive-expressive dynamic psychotherapy: A preliminary report. *Psychotherapy Research*, 6(2), 81–94.
- Barber, J. B., Crits-Christoph, P., & Luborsky, L. (1996). Effects of therapist adherence and competence on patient outcome in brief dynamic therapy. *Journal of Consulting and Clinical Psychology*, 3, 619–622.
- Beck, A. T., Rush, A. J., Shaw, B. F., Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford Press.
- Blackburn, I. M., James, I. A., Milne, D. L., Baker, C., Standart, S., Garland, A., & Reichelt, F. K. (2001). The revised cognitive therapy scale (CTS-R): Psychometric properties. *Behavioural and Cognitive Psychotherapy*, 29, 431–446.
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24, 385–396.
- Cohen, S., Mermelstein, R., Kamarck, T., & Hoberman, H. M. (1985). Measuring the functional components of social support. In I. G. Sarason & B. R. Sarason (Eds.), *Social support, theory, research, & applications*. Boston: Martenus Nijhoff & Dordrecht.
- DeRubeis, R. J., & Feeley, M. (1990). Determinants of change in cognitive therapy for depression. *Cognitive Therapy and Research*, 14(5), 469–482.
- Dixon, L., McFarlane, W. R., Lefley, H., Lucksted, A., Cohen, M., Falloon, I., Mueser, K., Miklowitz, K., Solomon, P., & Sondheim, D. (2001). Evidenced-based practices for services to families of people with psychiatric disabilities. *Psychiatric Services*, 52(7), 903–910.
- Dobson, K. S., Shaw, B. F., & Vallis, T. M. (1985). Reliability of a measure of the quality of cognitive therapy. *British Journal of Clinical Psychology*, 24, 295–300.
- Dyck, D. G., Hendryx, M. S., Short, R. A., Voss, W. D., & McFarlane, W. R. (2002). Service use among patients with schizophrenia in psycho-educational multiple-family group treatment. *Psychiatric Services*, 53(6), 749–754.
- Dyck, D. G., Short, R. A., Hendryx, M. S., Norell, D., Myers, M., Patterson, T., McDonell, M. G., Voss, W. D., & McFarlane, W. R. (2000). Management of negative symptoms among patients with schizophrenia attending multiple-family groups. *Psychiatric Services*, 51(4), 513–519.
- Feeley, M., DeRubeis, R. J., & Gelfand, L. A. (1999). The temporal relation of adherence to symptoms change in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, 67(4), 578–582.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1995). *Structured clinical interview for DSM-IV axis-I disorders*. New York: Columbia University.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378–382.
- Garety, P. A., Fowler, D., & Kuipers, E. (2000). Cognitive-behavioral therapy for medication-resistant symptoms. *Schizophrenia Bulletin*, 26, 73–85.
- Haddock, G., Devane, S., Bradshaw, T., McGovern, J., Tarrier, N., Kinderman, P., Baguley, I., Lacashire, S., & Harris, N. (2001). An investigation into the psychometric properties of the cognitive therapy scale for psychosis (CTS-Psy). *Behavioural and Cognitive Psychotherapy*, 29, 221–233.
- Hazel, N. A., McDonell, M. G., Short, R. A., Berry, C. M., Rodgers, M., Voss, W. D., & Dyck, D. G. (2004). Impact of multiple-family groups for outpatients with schizophrenia on caregivers' distress and resources. *Psychiatric Services*, 55(1), 35–41.
- Henry, W. P., Schacht, T. E., Strupp, H. H., Butler, S. F., & Binder, J. L. (1993). Effects of training in time-limited dynamic psychotherapy: Mediators of therapists' response to training. *Journal of Consulting and Clinical Psychology*, 61(3), 441–447.
- Hill, C. E., O'Grady, K. E., & Elkin, I. (1992). Applying the collaborative study psychotherapy rating scale to rate therapist adherence in cognitive-behavior therapy, interpersonal therapy, and clinical management. *Journal of Consulting and Clinical Psychology*, 60(1), 73–79.
- Hoglund, P., & Piper, W. E. (1995). Focal adherence in brief dynamic psychotherapy: A comparison of findings from two independent studies. *Psychotherapy*, 32(4), 618–628.

- Huey, S. J., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of change in multisystemic therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of Consulting and Clinical Psychology, 68*(3), 451–467.
- Huppert, J. D., Bufka, L. F., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2001). Therapists, therapist variables, and cognitive-behavioral therapy outcome in a multicenter trial for panic disorder. *Journal of Consulting and Clinical Psychology, 69*(5), 747–755.
- James, I. A., Blackburn, I. M., Milne, D. L., & Reichfelt, F. K. (2001). Moderators of trainee therapists' competence in cognitive therapy. *British Journal of Clinical Psychology, 40*, 131–141.
- Kavanagh, D. J. (1992). Recent developments in expressed emotion and schizophrenia. *British Journal of Psychiatry, 160*, 601–620.
- Kingdon, D. G., & Turkington, D. (2005). *Cognitive therapy of schizophrenia*. New York: Guilford Press.
- Lehman, A. F., & Steinwachs, D. M. (1998). At issue: Translating research into practice: The schizophrenia patient outcomes research team (PORT) treatment recommendations. *Schizophrenia Bulletin, 24*(1), 1–10.
- Luborsky, L., McLellan, T., Woody, G. E., O'Brien, C. P., & Auerbach, A. (1985). Therapist success and its determinants. *Archives of General Psychiatry, 42*, 602–611.
- Markowitz, J. C., Spielman, L. A., Scarvalone, P. A., & Perry, S. W. (2000). Psychotherapy adherence of therapists treating HIV-positive patients with depressive symptoms. *The Journal of Psychotherapy Practice and Research, 9*, 75–80.
- McDonell, M. G., Short, R. A., Hazel, N., Berry, C. M., & Dyck, D. G. (2006). Multiple family group treatment of outpatients with schizophrenia: Impact on service utilization. *Family Process, 45*, 359–373.
- McFarlane, W. R. (Ed.). (2002). *Multiple family groups in the treatment of severe psychiatric disorders*. New York: Guilford Press.
- McFarlane, W. R., Deakins, S. M., Gingerich, S. L., Dunne, E., Horen, B., & Newmark, M. (1991). *Multiple-family psycho-educational group treatment manual*. New York: Biosocial Treatment Division, New York State Psychiatric Institute.
- McFarlane, W. R., Link, B., Dushay, R., Marchal, J., & Crilly, J. (1995). Psycho-educational multiple family groups: Four-year relapse outcome in schizophrenia. *Family Process, 34*, 127–144.
- McFarlane, W. R., Lukens, E., Link, B., Dushay, R., Deakins, S. A., Newmark, M., Dunne, E. J., Horen, B., & Toran, J. (1995). Multiple-family groups and psycho-education in the treatment of schizophrenia. *Archives of General Psychiatry, 52*, 679–687.
- Multon, K. D., Kivlighan, D. M., & Gold, P. B. (1996). Changes in counselor adherence over the course of training. *Journal of Counseling Psychology, 41* (3), 356–365.
- O'Malley, S. S., Foley, S. H., Rounsaville, B. J., Watkins, J. T., Sotsky, S. M., & Imber, S. D. et al. (1988). Therapist competence and patient outcome in interpersonal psychotherapy of depression. *Journal of Consulting and Clinical Psychology, 56*(4), 496–501.
- Overall, J. E., & Gorham, D. R. (1962). The brief psychiatric rating scale. *Psychological Reports, 10*, 799–812.
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.
- Rounsaville, B. J., Chevron, E. S., Prusoff, B. A., Elkin, I., Imber, S., Sotsky, S. et al. (1987). The relations between specific and general dimensions of the psychotherapy process in interpersonal psychotherapy of depression. *Journal of Consulting and Clinical Psychology, 55*(3), 379–384.
- Sarason, I. G., Levine, H. M., Basham, R. B., & Sarason, B. R. (1983). Assessing social support: The social support questionnaire. *Journal of Personality and Social Psychology, 44*, 127–139.
- Schoenwald, S. K., Henggeler, S. W., Brondino, M. J., & Rowland, M. D. (2000). Multisystemic therapy: Monitoring treatment fidelity. *Family Process, 39*(1), 83–103.
- Shapiro, D. A., & Startup, M. (1992). Measuring therapist adherence in exploratory psychotherapy. *Psychotherapy Research, 2*(3), 193–203.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428.
- Spielberger, C. D., Gorsuch, R. L., & Lushene, R. (1970). *Manual for the State-Trait Anxiety Inventory: STAI ("Self-evaluation Questionnaire")*. Palo Alto: Consulting Psychologists Press.
- Spielberger C. D., Johnson E. H., Russell S. F., Crane R. J., Jacobs G. A., & Worden T. J. (1985). The experience and expression of anger: Construction and validation of an anger expression scale. In M. Chesney & R. Rosenman (Eds.), *Anger and hostility in cardiovascular and behavioral disorders* (pp. 6–30). New York: Hemisphere.
- Substance Abuse and Mental Health Administration. (2003). *Evidence-based practices: Shaping mental health services toward recovery*. Author.

- Svartberg, M. (1989). Manualization and competence monitoring of short-term anxiety-provoking psychotherapy. *Psychotherapy, 26*(4), 564–571.
- Turkington, D., Dudley, R., Warman, D. M., & Beck, A. T. (2004). Cognitive-behavioral therapy for schizophrenia: A review. *Journal of Psychiatric Practice, 10*, 5–16.
- Vaughn, C. E., Snyder, K. S., Jones, S., Freeman, W. B., & Falloon, I. R. H. (1984). Family factors in schizophrenic relapse. *Archives of General Psychiatry, 41*, 1169–1177.
- Vitaliano, P. P., Russo, J., Carr, J. E., Maiuro, R. D., & Becker, J. (1985). The ways of coping checklist: Revision and psychometric properties. *Multivariate Behavioral Research, 20*, 3–26.
- Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61*, 620–630.
- Weisman, A., Okazaki, S., Gregory, J., Goldstein, M. J., Thompson, M. C., Rea, M. M. et al. (1998). Evaluation of therapist competency and adherence to behavioral family management with bipolar patients. *Family Process, 37*, 107–121.