Journal of Counseling Psychology
2013, Vol. 60, No. 4, 601–609

0022-0167/13/$12.00 DOI: 10.1037/a0033539

BRIEF REPORT

Family Constellation Seminars Improve Psychological Functioning in a General Population Sample: Results of a Randomized Controlled Trial

Jan Weinhold, Christina Hunger, Annette Bornhäuser, Leoni Link, Justine Rochon, Beate Wild, and Jochen Schweitzer
University Hospital Heidelberg, Heidelberg, Germany

The study examined the efficacy of nonrecurring family constellation seminars on psychological health. We conducted a monocentric, single-blind, stratified, and balanced randomized controlled trial (RCT). After choosing their roles for participating in a family constellation seminar as either active participant (AP) or observing participant (OP), 208 adults ($M = 48$ years, $SD = 10$; 79% women) from the general population were randomly allocated to the intervention group (IG; 3-day family constellation seminar; 64 AP, 40 OP) or a wait-list control group (WLG; 64 AP, 40 OP). It was predicted that family constellation seminars would improve psychological functioning (Outcome Questionnaire OQ-45.2) at 2-week and 4-month follow-ups. In addition, we assessed the impact of family constellation seminars on psychological distress and motivational incongruence. The IG showed significantly improved psychological functioning ($d = 0.45$ at 2-week follow-up, $p = .003$; $d = 0.46$ at 4-month follow-up, $p = .003$). Results were confirmed for psychological distress and motivational incongruence. No adverse events were reported. This RCT provides evidence for the efficacy of family constellation in a nonclinical population. The implications of the findings are discussed.

Keywords: family constellation, treatment outcome, psychological functioning, psychological distress, motivational incongruence

Over the past 30 years, many forms of group-based psychosocial interventions—rooted in psychotherapy but defined as group counseling—have emerged (Corey, 2012): One popular but controversial method is the *family constellation*. A family constellation (FC) is a spatial arrangement of a family system in which individuals who are not members of the real family serve as *stand-ins* for the clients' family members (Schneider, 2009). Conceptually, the FC approach was influenced by group and family therapies (D. B. Cohen, 2006), especially the concept of transgenerational "invisible loyalties" between family members (Boszormenyi-Nagy & Spark, 1973). Technically, it integrates elements from psychodrama (Moreno, 1946) by using strangers as stand-ins, and from family sculptures (Duhl, Kantor, & Duhl, 1973) by setting up a spatial arrangement to symbolize a social system. One central aim is to identify patterns of transgenerational family dynamics: It is purported to visualize these family dynamics in a condensed and symbolic way. Typically, a series of FCs are conducted in a group setting in which approximately 25 participants, usually strangers, meet for a one-off, 3-day, facilitator-led seminar called a *family constellation seminar* (FCS). Of the 25 participants, about 15 will have signed up for a constellation of their own family (Schneider, 2009)—these are the active participants (APs)—whereas the rest are observing participants (OPs) who do not present their own FC but may represent members of the APs' families. Although an FCS resembles psychoeducational approaches such as marriage and relationship education (MRE; Hawkins, Blanchard, Baldwin, & Fawcett, 2008) in its brevity and interactional focus, it differs from MRE in that clients do not work with their actual partners. FCSs are not covered by the healthcare system, facilitators do not necessarily have a medical or psychotherapeutic background, and clients are not formally diagnosed. Because FCSs mainly address people seeking individual growth, or who may be experiencing temporarily maladjustment (DeLucia-Waack & Kalodner, 2005),

This article was published Online First August 19, 2013.

Jan Weinhold, Christina Hunger, Annette Bornhäuser, and Leoni Link, Institute of Medical Psychology, Center for Psychosocial Medicine, University Hospital Heidelberg, Heidelberg, Germany; Justine Rochon, Institute of Medical Biometry and Informatics, University Hospital Heidelberg; Beate Wild, Department of General Internal Medicine and Psychosomatics, Centre for Psychosocial Medicine, University Hospital Heidelberg; Jochen Schweitzer, Institute of Medical Psychology, Center for Psychosocial Medicine, University Hospital Heidelberg.

Jan Weinhold and Christina Hunger share first authorship for this article. All authors have agreed to authorship in the indicated order. There has been no prior publication, or the nature of any prior publication, and no financial interest in the research. Publication of this article has been made possible by the generous funding of the DFG (German Research Foundation) within the framework of the Collaborative Research Centre (SFB) 619 "Dynamics of Ritual," University of Heidelberg. We thank Diana Drexler and Gunthard Weber for conducting the family constellation seminars of this study.

Correspondence concerning this article should be addressed to Jan Weinhold, Institute of Medical Psychology, Center for Psychosocial Medicine, University Hospital Heidelberg, Bergheimer Straße 20, D-69115 Heidelberg, Germany. E-mail: jan.weinhold@med.uni-heidelberg.de

we consider them a form of brief, solution-focused group counseling.

The procedure for each individual FC is as follows (Schneider, 2009): First, the facilitator briefly interviews one AP about the issue at stake and his or her family history. A decision is made about which system will be set up spatially (e.g., the family of origin or the current family). Next, the AP selects group members to act as stand-ins for family members including him- or herself. Then, the AP places all stand-ins in a defined, clear space. Spatial distances, angles, and body postures are meant to correspond to the APs' inner image of the family system in question. Once this initial constellation is set up, the AP takes a seat to observe the process. Initially, the stand-ins do not move, interact, or speak. When cued by the facilitator, they voice perceptions and observations on the basis of their position in the constellation (e.g., "I don't feel good" or "a good place") and remark on interpersonal aspects relating to the others (e.g., "My ex-husband is between me and my new partner"). On the basis of these statements, the facilitator rearranges the constellation according to the theoretical principles of FC, such as belonging and generational order (Sparrer, 2006). A "solution picture" denotes the final constellation that is perceived by the AP to have changed for the better.

Empirical studies of FCSs are rare and use small samples. Although a systematic review identified 19 studies that explicitly investigated the efficacy of FCSs (Reinhard, 2012), most have not been published in scientific journals and show considerable methodological shortcomings. A quasi-experimental study with fewer methodological deficits indicated higher self-esteem and reduced psychopathological symptoms in a nonclinical population 4 months after an FCS (Höppner, 2001). Nor has the efficacy of FCSs been properly assessed—and this may prove difficult. First, the qualifications of FC facilitators vary considerably. Second, FCSs are emotionally intense, one-off interventions not typically integrated into an ongoing counseling process. Furthermore, the intervention also attracts clients who may have clinical problems. Finally, the approach has been criticized for some of its theoretical assumptions, for example, that the symbolic constellation represents facts of the family history (Haas, 2009). Although FCSs seem to have the potential to facilitate personal growth and improve psychological functioning, existing results are heterogeneous and methodologically questionable, pointing to the need for a methodologically sound study. Because FCSs are commonly conducted in nonclinical settings, we deemed it appropriate to select wide outcome measures on the basis of general models of psychotherapeutic and counseling change processes; our outcome measures capture both clinical and nonclinical specifications. Lambert, Hill, Bergin, and Garfield (1994) postulated psychological functioning and psychological distress as central outcomes in psychotherapy and counseling. Within the framework of general psychotherapy, *motivational incongruence*, defined as "the unsatisfactory realization of . . . motivational goals" (Grosse Holtforth & Grawe, 2003, p. 315), correlates with the severity of psychological disorders and reduced well-being; therefore, reducing motivational incongruence is an overarching goal in counseling or psychotherapy (Grosse Holtforth & Grawe, 2003). We hypothesized that a 3-day FCS would significantly improve psychological functioning at 2-week and 4-month follow-ups. In ancillary analyses, we explored whether psychological distress and motivational incongruence had decreased at both follow-ups. We also expected that APs would benefit more from the intervention than OPs: Because their own problems are actively addressed, they receive more specific attention and care. Finally, we assessed potential harms of the intervention.

Method

Participants

In order to contribute to the external validity by designing FCSs in our study similar to those carried out in naturalistic settings, the inclusion criteria were minimal. Persons of either gender age 18 or older could take part if they (a) participated in a 3-day FCS, (b) made their own decision to participate as AP or OP, (c) accepted random allocation to the intervention group (IG) or wait-list control group (WLG), (d) agreed to randomized assignment to one of the two study facilitators, and (e) would abstain from taking part in another FCS until completion of the study. To avoid harm, participants were excluded from the intervention if they showed (a) acute suicidal tendencies, (b) an acute psychotic episode, or (c) an acute drug or alcohol intoxication. All participants gave their informed consent. Study participants were recruited from a general population via flyers placed in target contexts and a website created for advertising. Potential study participants who contacted the study center after public recruitment were screened according to our eligibility criteria, and the conditions of participation and randomization were explained; they were also sent information about the aim of the study, a consent form, and a pretest (biographical) questionnaire. Upon their informed consent, they were included in the study and randomly assigned to either the IG or the WLG. Recruitment ended when we reached 208 participants. This study was approved by the Ethics Committee of the Heidelberg Medical Faculty (S-178/2011) and registered with Clinical Trials (www.clinicaltrials.gov; Nr. NCT 01352325).

Sample characteristics. For an overview of participant flow, see Figure 1. None of the 208 study participants dropped out. Five IG participants (2.4%) did not receive the treatment, three IG participants (1.4%) and four WLG participants (1.9%) committed major protocol deviations by attending another FCS during the study period. Two IG participants (1%) and one WLG participant (0.5%) failed to complete the primary outcome instrument at the 4-month follow-up. Intention-to-treat analyses involved all 208 participants. Per-protocol analysis was conducted for 193 participants (IG: 94, WLG: 99; 92.8%). The two study arms were balanced with respect to baseline characteristics (see Table 1).

Study facilitators. To ensure the quality of the treatment and the competency to deal with study participants who were potentially in the clinical range, eligibility criteria for the facilitators were to (a) be either a licensed psychiatrist or a psychotherapist, (b) have at least 20 years of professional experience as a psychological therapist, and (c) have at least 10 years of professional experience with FCSs. Facilitator 1 was a 71-year-old male licensed psychiatrist and psychotherapist with 30 years' experience conducting FCSs; Facilitator 2 was a 53-year-old female clinical psychologist and licensed psychotherapist with 20 years' FCS experience.

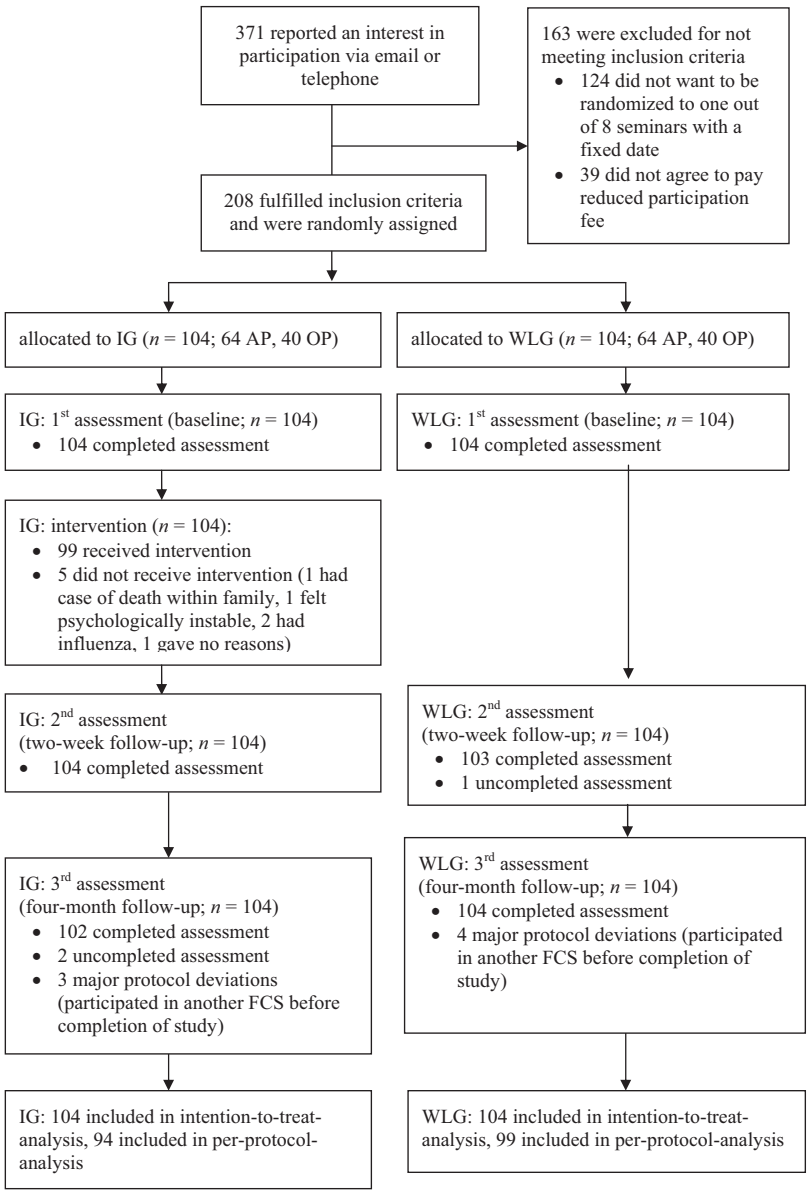

Figure 1. Participant flow. IG = intervention group; WLG = wait-list control group; AP = active participants; OP = observing participants; FCS = family constellation seminar.

Study Design

The study was designed as a prospective, monocentric, parallel-group, randomized controlled trial (RCT). After stratification by role (voluntary choice to participate either as AP or OP prior to randomization), participants were randomly assigned in a 1:1 ratio to attend a 3-day FCS, either in the IG (FCSs 1–4) or in the WLG (FCSs 5–8), which received the same treatment 4 months after the IG. Each of the two study facilitators conducted four FCSs (Facilitator 1: FCS 1, 2, 5, and 6; Facilitator 2: FCS 3, 4, 7, and 8), and participants decided whether to take part in the study as AP or OP. The treatment outcome was measured at three assessment points: baseline, 2-week follow-up, and 4-month follow-up.

Sample Size and Randomization

A power analysis for repeated measures mixed-design analysis of variance (ANOVA) using the software G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) revealed that a total sample size of 208 participants (4 FCSs × 26 participants = 104 participants per study arm) was sufficient to detect small effect sizes of Cohen's $f = 0.1$ (J. Cohen, 1992) with a power of about 90% at a significance level of .05. We used restricted randomization to obtain a balanced sample size between groups for each seminar. The random allocation rule was implemented according to the restricted shuffled approach (Schulz & Grimes, 2002). The randomization procedure carried out someone not otherwise involved in the study.

Table 1
Sample Characteristics

Variable	IG (n = 104)			WLG (n = 104)		
	Total	AP (n = 64)	OP (n = 40)	Total	AP (n = 64)	OP (n = 40)
Age (years), mean (*SD*)	47 (9)	47 (9)	48 (10)	48 (10)	48 (10)	49 (10)
Gender: Female, *n* (%)	87 (84%)	51 (80%)	36 (90%)	77 (74%)	45 (70%)	32 (80%)
Marital status: Married or in partnership	69 (66%)	42 (66%)	27 (68%)	77 (74%)	46 (72%)	31 (78%)
Education: High school diploma	92 (88%)	57 (89%)	35 (88%)	89 (86%)	40 (63%)	39 (98%)
Professional status: Employed	98 (94%)	59 (92%)	39 (98%)	94 (90%)	55 (86%)	39 (98%)
Helping professions	65 (63%)	38 (59%)	27 (68%)	54 (52%)	31 (48%)	23 (58%)
Nationality: German	101 (97%)	63 (98%)	38 (95%)	99 (95%)	61 (95%)	38 (95%)
Previous experience with FC	83 (80%)	51 (80%)	32 (80%)	82 (79%)	54 (84%)	28 (70%)

Note. IG = intervention group; WLG = wait-list control group; AP = active participants; OP = observing participants; FC = family constellation.

Neither the study participants nor the two study facilitators were informed about the parallel design of the study, although outcome assessors were aware of the study design and of which study arm each participant was allocated to.

Intervention

The intervention consisted of a 3-day FCS. All FCSs took place in a group therapy room at the Institute of Medical Psychology at University Hospital Heidelberg, Heidelberg, Germany.

Standardization of treatment and assessment of treatment integrity. To standardize the conduct of FCSs in our study, we developed a manual based on previous observations of FCSs consisting of 11 obligatory procedures of the FCS and seven obligatory procedures of each individual FC, including an operational definition for each procedure. Researchers who would later observe the study FCSs were trained by observing additional FCSs conducted by both study facilitators. The manual is available from the first author (JW). Two members of the research team observed each FCS of the IG in its entirety, including all individual FCs. They independently rated all treatment procedures according to our FCS manual by checking off each procedure if fulfilled. Interrater reliability of these ratings was computed as a percentage of the agreement between the two raters. A more robust measure of interrater reliability (e.g., Cohen's kappa) could not be applied due to the low degree of total observer variance. Interrater reliability between the two raters ranged between 96.5% and 97.3% in the four FCSs in the IG. Overall treatment integrity for each FCS was calculated as percentage of treatment components defined by the FCS manual across sessions that were implemented as planned (Perepletchikova, 2011). As a measure for the implementation, we used the more conservative ratings of the two raters. Overall treatment integrity for the four FCSs within the IG ranged between 96.5% and 100%.

Assessment Instruments

Psychological functioning. The Outcome Questionnaire (OQ-45.2; Lambert, Hannöver, Nisslmüller, Richard, & Kordy, 2002; Lambert et al., 2004) was used to assess psychological functioning. The OQ-45.2 is a self-reporting instrument that measures psychotherapeutic change over the previous week. Its total score (OQ-TOT) indicates symptom distress, quality of interpersonal relations, and social role performance. The OQ-45.2 features test–retest reliability of .84, excellent internal consistency at .93, sensitivity to change, and concurrent validity with a variety of self-report scales (Haug, Puschner, Lambert, & Kordy, 2004; Lambert et al., 2004; Vermeersch et al., 2004). In the present study, Cronbach's alpha for the OQ-TOT at baseline was .91.

Psychological distress. The Questionnaire for the Evaluation of Treatment Progress (FEP; Lutz et al., 2009) was used for measuring psychological distress. The FEP measures how persons felt over the previous week. Its total score (FEP-TOT) indicates well-being, symptom distress, interpersonal relationships, and congruence. The FEP shows test–retest reliability of .77 (1 week) and internal consistency of .94. Concurrent validity and sensitivity to change were demonstrated (Lutz et al., 2009). In the present study, Cronbach's alpha for the FEP-TOT at baseline was .95.

Motivational incongruence. We used the short form of the Incongruence Questionnaire (INK-SF; Grosse Holtforth, Grawe, & Tamcan, 2004) to measure motivational incongruence. Its total score (INK-SF-TOT) indicates approach goals (the maximization of desirable outcomes) and avoidance goals (the minimization of unwanted outcomes). One-week test–retest reliability was .81, and internal consistencies ranged from .75 to .91. Criterion validity was assessed using established instruments for psychopathological symptoms and quality of life (Grosse Holtforth et al., 2004). In the present study, Cronbach's alpha for the INK-SF-TOT at baseline was .88.

Adverse events. Harms were assessed by means of passive surveillance (Ioannidis et al., 2004). Study participants were verbally instructed by members of the research team to contact them personally if they experienced any *adverse events*, defined as any psychological or somatic problems that they felt may be attributable to the intervention at any time between the beginning of their FCS and the end of the study.

Statistical Analyses

All analyses were calculated using the IBM SPSS statistical package (Version 19.0). Continuous variables are expressed as means with standard deviations. Categorical variables are presented as absolute numbers and percentages.

Clustering effects. Because participants within groups (i.e., FCSs) may be more similar to each other than participants in different FCSs, a two-level linear regression analysis was per-

formed to account for the clustering effects at higher levels (i.e., participants nested within FCSs). The intraclass correlation (ICC) coefficient from the random intercept model with participants (Level 1) and FCS (Level 2) was calculated for all outcome measures (Kreft & de Leeuw, 1998). An ICC greater than zero indicates the existence of clustering effects: Either the standard statistical analyses must be adjusted for these effects or multilevel analyses that account for such dependence of the data should be conducted.

Facilitator effects. In order to examine potential facilitator effects, 2 (facilitator) × 3 (time) mixed-design ANOVAs were conducted for the three outcome measures, with facilitator as the independent variable and outcome measure as the dependent variable (OQ-45-TOT, FEP-TOT, INK-SF-TOT).

Treatment outcome. All statistical tests were conducted two-tailed. Single missing values of less than 20% for each scale of all outcome measures were replaced with the conditional mean values of the four subgroups (APs in IG and WLG; OPs in IG and WLG). First, intention-to-treat analyses with all 208 randomized participants were conducted, followed by per-protocol analysis, including all participants without major protocol deviations. Mixed-design ANOVAs were performed to identify differences between IG and WLG. For these ANOVAs, the factors were group (IG, WLG), participant status (AP, OP), and time (baseline, 2-week, and 4-month follow-up). In the case of significant interactions, simple effects analyses within and between groups were performed (Howell, 2002). Because treatment outcome was measured at three assessment points, within-group effects were further analyzed by comparisons between baseline and 2-week follow-up (Contrast A) and by comparisons between 2-week and 4-month follow-ups (Contrast B). To identify differences in treatment outcome according to participant status, ancillary mixed-design ANOVAs were performed for all outcome measures, excluding the WLG. For these ANOVAs, the factors were participant status within the IG (AP, OP) and time (baseline, 2-week, and 4-month follow-up). Effect sizes are presented as partial eta-squared values (η^2) and Cohen's d, calculated as the difference between the means divided by the pooled standard deviation $\left(d = \frac{x_1^2 - x_2^2}{\sqrt{s_1^2 + s_2^2/2}}\right)$. Classification of effect sizes was as follows: $\eta^2 \geq .010$, small effect; $\eta^2 \geq .060$, medium effect; $\eta^2 \geq .140$, large effect; Cohen's $d \geq 0.20$, small effect; $d \geq 0.50$, medium effect; $d \geq 0.80$, large effect (J. Cohen, 1988).

Clinical significance. We assessed reliably significant change with the reliable change index (RCI; Jacobson & Truax, 1991). We used previously established internal consistency reliabilities of the outcome measures to calculate the RCI (Bauer, Lambert, & Nielsen, 2004). This assessment estimates the direction and amount of reliable change (positive direction, negative direction, no change). To assess clinically meaningful change, we calculated a cutoff score for each outcome measure representing an estimated difference between "distressed" and "normal" of the respective outcome. We used Jacobson and Truax's (1991) cutoff C, the weighted average between a clinical sample and a nonclinical sample. Clinical and nonclinical norms for all outcome measures were determined on the basis of previous publications (Grosse Holtforth et al., 2004; Lambert et al., 2002; Lutz et al., 2009).

Results

Group Effects

Clustering. For all outcome measures, the corresponding ICC coefficient at Level 2 (FCS) was zero. Therefore, clustering effects were not expected, and results from standard statistical analyses (mixed-design ANOVAs) are presented.

Facilitators. For all outcome measures, the Time × Facilitator interaction effect was not significant, indicating no differences in the degree of change based on the two different facilitators. Additionally, we analyzed a graph of average change for each outcome over three time points for each facilitator: For both, the pattern was similar across outcome variables.

Treatment Outcome

Results for primary and secondary outcome measures are displayed in Table 2. Due to the nonsignificance of the three-way interaction (Group × Time × Participant status) for all outcome measures, we analyzed only the Time × Group and the Time × Participant status interaction by simple effects analyses.

Primary outcome: Psychological functioning. For the OQ-TOT, mixed-design ANOVA showed that there was a significant but small Time × Group interaction effect in the intention-to-treat analysis (see Table 2). The between-group simple effects analysis indicates that the IG showed significantly stronger improvement of psychological functioning compared with the WLG at both follow-ups (see Figure 2). Effect sizes were small at both follow-up assessment points. Contrast A revealed significance comparing baseline with 2-week follow-up, $F(1, 204) = 23.55$, $p = .000$, $\eta^2 = .104$. Contrast B was not statistically significant. Within-group simple effects analysis demonstrated that this effect was due to a significant improvement over time in the IG, $F(1.73, 176.84) = 17.85$, $p = .000$, $\eta^2 = .149$. Participants in the WLG demonstrated no significant improvement over time. Thus, differences between the two study arms are attributable to an improvement of psychological functioning in the participants of the IG at 2-week follow-up—improvements that remained stable over the following 4 months. Per-protocol analysis supported the results of this intention-to-treat analysis.

Secondary outcomes: Psychological distress. For the FEP-TOT, mixed-design ANOVA demonstrated a significant Time × Group interaction effect with a small effect size. Between-group simple effects analysis indicated that the IG reported significantly decreased psychological distress compared with the WLG at both follow-ups, with medium effect sizes. Contrast A revealed significance comparing baseline with 2-week follow-up, $F(1, 204) = 19.91$, $p = .000$, $\eta^2 = .089$. Contrast B was not statistically significant. Within-group simple effects analysis showed that this effect was due to significant improvement over time in the IG, $F(1.72, 175.27) = 14.27$, $p = .000$, $\eta^2 = .123$. Participants in the WLG showed no significant improvement over time. Thus, differences between the IG and the WLG in psychological distress are attributable to changes reported by the IG at 2-week follow-up.

Motivational incongruence. For the INK-SF-TOT, mixed-design ANOVA showed a significant but small Time × Group interaction effect. Between-group simple effects analysis indicated that the IG reported decreased motivational incongruence com-

Table 2
Primary and Secondary Outcome: Descriptive Data at Baseline, 2-Week and 4-Month Follow-Up for IG and WLG, ANOVA, Between-Group Simple Effects Analyses, and Effect Sizes Between Groups at 2-Week and 4-Month Follow-Up

Variable	Group	Baseline	2-week	4-month	ANOVA interaction: Time × Group[a]	Simple effects on the between-subjects factor (IG/WLG) 2-week follow-up[a]			Simple effects on the between-subjects factor (IG/WLG) 4-month follow-up[a]		
		M (*SD*)	*M* (*SD*)	*M* (*SD*)		*F* (*df*)	*p*	*d*	*F* (*df*)	*p*	*d*
OQ-45-TOT	FCS (n = 104)	49.94 (17.78)	42.63 (17.39)	42.46 (17.40)	$F(1.83, 373.07) = 12.89$, $p = .000$, $\eta^2 = .059$	9.35 (1, 204)	.003	0.45	9.22 (1, 204)	.003	0.46
	WLG (n = 104)	49.91 (18.71)	51.03 (20.22)	50.79 (18.83)							
FEP-TOT	FCS (n = 104)	2.15 (0.56)	1.92 (0.51)	1.92 (0.53)	$F(1.82, 371.37) = 9.81$, $p = .000$, $\eta^2 = .046$	12.82 (1, 204)	.000	0.51	11.24 (1, 204)	.001	0.51
	WLG (n = 104)	2.16 (0.57)	2.21 (0.62)	2.20 (0.57)							
INK-SF-TOT	FCS (n = 104)	2.21 (0.56)	2.02 (0.59)	2.03 (0.57)	$F(1.92, 390.65) = 8.33$, $p = .000$, $\eta^2 = .039$	14.78 (1, 204)	.000	0.55	12.94 (1, 204)	.000	0.52
	WLG (n = 104)	2.29 (0.55)	2.35 (0.62)	2.33 (0.59)							

Note. IG = intervention group; WLG = wait-list control group; ANOVA = analysis of variance; OQ-45 = Outcome Questionnaire; TOT = total; FEP = Questionnaire for the Evaluation of Treatment Progress; INK-SF = short form of the Incongruence Questionaire; FCS = family constellation seminar.
[a] Greenhouse–Geisser-corrected values.

Figure 2. Outcome Questionnaire total score (OQ-45-TOT). IG = intervention group; WLG = wait-list control group.

pared with the WLG. Effect sizes were medium at both follow-ups. Contrast A revealed significance comparing baseline with 2-week follow-up, $F(1, 204) = 18.12$, $p = .000$, $\eta^2 = .082$. Contrast B was not significant. Within-group simple effects analysis demonstrated that this effect is based on a significant improvement over time in the IG, $F(1.86, 189.74) = 10.12$, $p = .000$, $\eta^2 = .090$. No significant changes over time were observed within the WLG. Thus, differences between the IG and the WLG in motivational incongruence are attributable to changes reported by the IG at 2-week follow-up.

Treatment outcome according to participant status. Comparisons of outcomes according to participant status (AP vs. OP within the IG), excluding the WLG, detected no significant differences between the two subgroups. Mixed-design ANOVAs for the total scores of all outcome measures revealed no significant Time × Participant status interactions; the main effects for participant status were also not significant.

Clinical significance. The categorization of participants of both study arms at baseline, 2-week, and 4-month follow-up, including RCI and cutoff for each outcome measure, appear in Table 3. At both follow-ups, the majority of participants in the IG did not change reliably in either of the outcome measures. Of the participants, 28% in the IG attained a reliable change below the cutoff in psychological functioning (OQ-45.2) at 2-week follow-up and 25% at 4-month follow-up. The percentages of RCI below the cutoffs for psychological distress (FEP) and motivational incongruence (INK-SF) were similar. Between 2% (INK-SF at 2-week follow-up) and 12% (FEP at 4-month follow-up) of the participants in the IG deteriorated. In the WLG, fewer participants recovered (reliable change below the cutoff) and more participants deteriorated compared with the IG (see Table 3).

Adverse effects. Throughout the entire study, neither during nor after the FCS did any study participant report adverse events to the research team.

Table 3
Primary and Secondary Outcomes: Clinically Significant Change for IG and WLG

Variable	Outcome	*n*	Below cutoff *n* (%)	Recovered *n* (%)	Improved *n* (%)	Unchanged *n* (%)	Deteriorated *n* (%)
IG							
Baseline	OQ-45	104	77 (74)				
	FEP	104	74 (71)				
	INK-SF	104	64 (62)				
2-week follow-up	OQ-45	104		29 (28)	0 (0)	72 (69)	3 (3)
	FEP	104		29 (28)	1 (1)	68 (65)	6 (6)
	INK-SF	104		25 (24)	1 (1)	76 (73)	2 (2)
4-month follow-up	OQ-45	104		26 (25)	1 (1)	73 (70)	4 (4)
	FEP	104		34 (33)	3 (3)	55 (53)	12 (12)
	INK-SF	104		25 (24)	2 (2)	67 (64)	10 (10)
WLG							
Baseline	OQ-45	104	79 (76)				
	FEP	104	72 (69)				
	INK-SF	104	63 (61)				
2-week follow-up	OQ-45	104		6 (6)	0 (0)	91 (88)	7 (7)
	FEP	104		7 (7)	2 (2)	75 (72)	20 (19)
	INK-SF	104		2 (2)	1 (1)	93 (89)	8 (8)
4-month follow-up	OQ-45	104		8 (8)	1 (1)	82 (79)	13 (13)
	FEP	104		20 (19)	1 (1)	60 (58)	23 (22)
	INK-SF	104		10 (10)	2 (2)	78 (75)	14 (13)

Note. IG = intervention group; WLG = wait-list control group; Below cutoff = within functional range; Recovered = reliable change below the cutoff; Improved = reliable change not below the cutoff; Unchanged = no reliable change; Deteriorated = reliable change in a negative direction; OQ-45 = Outcome Questionnaire; FEP = Questionnaire for the Evaluation of Treatment Progress; INK-SF = short form of the Incongruence Questionnaire. Reliable change index: OQ-45: ±13.58; FEP: ±0.33; INK-SF: ±0.20; Cutoff: OQ-45: 63.00; FEP: 2.45; INK-SF: 2.40.

Discussion

In this RCT study, we investigated the efficacy of FCSs with regard to variables of psychological health. Results indicate improvements in psychological functioning following an FCS at 2-week and 4-month follow-ups, compared with a wait-list condition. Exploratory analyses showed reduced psychological distress and reduced motivational incongruence in the IG at both follow-ups. Small to medium effect sizes were demonstrated for the three outcome measures. Effect sizes at 4-month follow-up indicated sustainable effects. Effect sizes in our study were smaller than those achieved with multisession psychotherapy (Lambert & Ogles, 2004) or measured in other counseling interventions (Minami et al., 2009). They do correspond, however, to those of the interpersonal psychoeducational groups often conducted as weekend seminars, for example, MRE (Hawkins et al., 2008). Although our outcome measures cover clinical and nonclinical gradations, their specificity for assessing FCSs has its limits. Two other studies that investigated goal attainment (Bornhäuser et al., 2013) and experience in personal social systems (Hunger, Bornhäuser, Link, Schweitzer, & Weinhold, 2013) yielded higher effect sizes than the outcome measures reported here.

The clinical significance of our results should be interpreted modestly. The mean change for psychological functioning in the IG was smaller than the RCI of the OQ-45.2. Although the majority of study participants in the IG did not change reliably at 2-week and 4-month follow-ups, given that the majority of them were in the functional range at baseline, the proportions of clinically significant change in our study are comparable to those reported in the literature (e.g., Lambert & Ogles, 2004). Results regarding clinical significance were consistently better in the IG than in the WLG at both follow-up assessment points, supporting the efficacy of FCSs. However, few participants in the IG deteriorated at both follow-ups. We hypothesize that such deterioration may be attributed to the fact that FCSs are experientially intense single-session interventions, which are not integrated into an ongoing counseling process.

APs and OPs within the IG did not differ in treatment outcomes. There may be group factors from which both subgroups benefit equally (cohesion, group climate, alliance, empathy; Johnson, Burlingame, Olsen, Davies, & Gleave, 2005). During all FCSs, we observed the development of a positive group climate and strong cohesion forming between all group members. OPs experienced many FCs with strong emotional and existential issues, which led to a high degree of empathy. By serving as stand-ins, they vicariously experienced symbolized family dynamics. The two experienced study facilitators managed to establish high-quality working relationships and alliances with all group members. Thus, the difference between "active" and "observing" participants may be much less important than we expected.

The major strength of this study is that it is the first RCT to investigate the efficacy of FCSs relating to psychological functioning. We successfully implemented FCSs as conducted in naturalistic settings. The participants' voluntary choice of status (AP or OP), adherence to the prevailing procedures of typical FCSs, and the absence of research-triggered disturbances contributed to the external validity. Interventions were performed with excellent treatment integrity. The attrition rate was very low. No study participants dropped out of the study before completion, and 92.8% of them were included in per-protocol analyses.

Our study has some limitations: We used a wait-list control design instead of an appropriate control group treatment (Rifkin, 2007). Although effect sizes were stable after 4 months, changes

within the IG cannot be attributed to specific characteristics of the intervention. Future studies should compare FCSs with other family-related interventions (e.g., multifamily counseling or psychodrama seminars). Due to the brevity of the intervention and the small effect sizes found, FCs should be examined as an add-on to established and more extensive interventions (e.g., family counseling) to detect specific effects of the method.

The self-selected sample also limits generalizability. Study participants responded to advertisements and were mainly female, well-educated, employed (often in helping professions), and had previous experience with FCSs. Because the informed consent stated that FCSs may contribute to the improvement of psychological health, treatment expectations may have arisen. Provided that participants were attracted to our study because of their previous experience with FCSs, positive expectations about the intervention may have influenced results. Helping professionals may generally be predisposed to experience positive effects of FCSs. Finally, it would have been preferable to ensure a balanced gender ratio. We selected only two study facilitators: Their long-term experience and expertise in conducting FCSs also limits generalizability. Their allegiance to the treatment, a factor known to contribute to the outcome regardless of the specific intervention (Leykin & DeRubeis, 2009), was high. Therefore, our results can be attributed in part to facilitator allegiance. In future studies, a broader range of facilitators with various professional backgrounds and degrees of FCS experience should be used. Although investigators were not proponents of the FC approach, five of the seven authors are systemic psychotherapists; thus, their presence as researchers in the FCSs could also have influenced the group processes and the outcomes. Finally, the lack of a standardized diagnostic procedure may be considered a shortcoming. Future research should replicate our results with clinical populations.

We boldly conclude that our results point to a potentially positive impact of FCSs regarding psychological health. Despite the limitations of this first RCT, the results encourage further evaluations with refined study designs.

References

Bauer, S., Lambert, M. J., & Nielsen, S. L. (2004). Clinical significance methods: A comparison of statistical techniques. *Journal of Personality Assessment, 82,* 60–70. doi:10.1207/s15327752jpa8201_11

Bornhäuser, A., Hunger, C., Link, L., Wolff, E., Schweitzer, J., & Weinhold, J. (2013). *Randomized controlled trial on Family Constellation Seminars: Findings on goal content, goal incidence, and goal attainment*. Manuscript submitted for publication.

Boszormenyi-Nagy, I., & Spark, G. (1973). *Invisible loyalties: Reciprocity in intergenerational family therapy*. New York, NY: Harper & Row.

Cohen, D. B. (2006). "Family constellations": An innovative systemic phenomenological group process from Germany. *Family Journal, 14,* 226–233. doi:10.1177/1066480706287279

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. doi:10.1037/0033-2909.112.1.155

Corey, G. (2012). *Theory and practice of group counseling* (8th ed.). Monterey, CA: Brooks/Cole.

DeLucia-Waack, J. L., & Kalodner, C. R. (2005). Contemporary issues in group practice. In S. A. Wheelan (Ed.), *The handbook of group research and practice* (pp. 65–84). Thousand Oaks, CA: Sage. doi:10.4135/9781412990165.d6

Duhl, F., Kantor, D., & Duhl, B. (1973). Learning, space and action. A primer of family sculpture. In D. Bloch (Ed.), *Techniques of family psychotherapy* (pp. 47–63). New York, NY: Grune & Stratton.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149–1160. doi:10.3758/BRM.41.4.1149

Grosse Holtforth, M., & Grawe, K. (2003). The incongruence questionnaire (INK). An instrument for the analysis of motivational incongruence. *Zeitschrift für Klinische Psychologie und Psychotherapie: Forschung und Praxis, 32,* 315–323. doi:10.1026//1616-3443.32.4.315

Grosse Holtforth, M., Grawe, K., & Tamcan, Ö. (2004). *Inkongruenzfragebogen: INK* [Incongruence Questionnaire: INK]. Göttingen, Germany: Hogrefe.

Haas, W. (2009). *Das Hellinger-Virus. Zu Risiken und Nebenwirkungen von Aufstellungen* [The Hellinger-Virus. About risks and side effects of Systemic Constellations] (Vol. 2). Kröning, Germany: Asanger Verlag.

Haug, S., Puschner, B., Lambert, M. J., & Kordy, H. (2004). Assessment of change in psychotherapy with the German version of the Outcome Questionnaire (OQ-45.2). *Zeitschrift für Differentielle und Diagnostische Psychologie, 25,* 141–151. doi:10.1024/0170-1789.25.3.141

Hawkins, A. J., Blanchard, V. L., Baldwin, S. A., & Fawcett, E. B. (2008). Does marriage and relationship education work? A meta-analytic study. *Journal of Consulting and Clinical Psychology, 76*(5), 723–734. doi:10.1037/a0012584

Höppner, G. (2001). *Heilt Demut - wo Schicksal wirkt? Evaluationsstudie zu Effekten des Familien-Stellens nach Bert Hellinger* [Does humility restore where destiny operates? Evaluation study on the effects of Family Constellations by Bert Hellinger]. München, Germany: Profil-Verlag.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury Thomson Learning.

Hunger, C., Bornhäuser, A., Link, L., Schweitzer, J., & Weinhold, J. (2013). *Improving experience in personal social systems through family constellation seminars: Results of a randomized controlled trial*. Manuscript submitted for publication.

Ioannidis, J. P. A., Evans, S. J. W., Gøtzsche, P. C., O'Neill, R. T., Altman, D. G., Schulz, K., & Moher, D. (2004). Better reporting of harms in randomized trials: An extension of the CONSORT statement. *Annals of Internal Medicine, 141,* 781–788. doi:10.7326/0003-4819-141-10-200411160-00009

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12–19. doi:10.1037/0022-006X.59.1.12

Johnson, J. E., Burlingame, G. M., Olsen, J. A., Davies, D. R., & Gleave, R. L. (2005). Group climate, cohesion, alliance, and empathy in group psychotherapy: Multilevel structural equation models. *Journal of Counseling Psychology, 52,* 310–321. doi:10.1037/0022-0167.52.3.310

Kreft, I., & de Leeuw, J. (1998). *Introducing multivlevel modeling*. Thousand Oaks, CA: Sage.

Lambert, M. J., Hannöver, W., Nisslmüller, K., Richard, M., & Kordy, H. (2002). Questionnaire on the results of psychotherapy: Reliability and validity of the German translation of the Outcome Questionnaire 45.2. *Zeitschrift für Klinische Psychologie und Psychotherapie: Forschung und Praxis, 31,* 40–46. doi:10.1026//1616-3443.31.1.40

Lambert, M. J., Hill, C. E., Bergin, A. E., & Garfield, S. L. (1994). Assessing psychotherapy outcomes and processes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 72–113). Oxford, England: Wiley.

Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, S. C., Hamilton, S., Reid, R. C., . . . Burlingame, G. B. (2004). *Administration and scoring manual for the OQ-45.2 (Outcome Questionnaire)* [Manual]. Orem, UT: American Professional Credentialing Services.

Lambert, M. J., & Ogles, B. M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 139–193). New York, NY: Wiley.

Leykin, Y., & DeRubeis, R. J. (2009). Allegiance in psychotherapy outcome research: Separating association from bias. *Clinical Psychology-Science and Practice, 16,* 54–65. doi:10.1111/j.1468-2850.2009.01143.x

Lutz, W., Schürch, E., Stulz, N., Böhnke, J. R., Schöttke, H., Rogner, J., & Wiedl, K. H. (2009). Entwicklung und psychometrische Kennwerte des Fragebogens zur Evaluation von Psychotherapieverläufen (FEP) [Development and psychometric properties of a questionnaire to evaluate psychotherapeutic processes]. *Diagnostica, 55*(2), 106–116. doi:10.1026/0012-1924.55.2.106

Minami, T., Davies, D. R., Tierney, S. C., Bettmann, J. E., McAward, S. M., Averill, L. A., . . . Wampold, B. E. (2009). Preliminary evidence on the effectiveness of psychological treatments delivered at a university counseling center. *Journal of Counseling Psychology, 56,* 309–320. doi:10.1037/a0015398

Moreno, J. L. (1946). *Psychodrama, first volume*. New York, NY: Beacon House. doi:10.1037/11506-000

Perepletchikova, F. (2011). On the topic of treatment tntegrity. *Clinical Psychology: Science and Practice, 18,* 148–153. doi:10.1111/j.1468-2850.2011.01246.x

Reinhard, A. (2012). *Systematische Übersicht über die Wirksamkeit von Systemaufstellungen*. [Systematic review of the efficacy of systemic constellations]. Unpublished manuscript, Division of Organizational Psychology in Medicine, Institute for Medical Psychology, University Hospital Heidelberg.

Rifkin, A. (2007). Randomized controlled trials and psychotherapy research. *American Journal of Psychiatry, 164,* 7–8. doi:10.1176/appi.ajp.164.1.7

Schneider, J. R. (2009). *Family constellations: Basic principles and procedures*. Heidelberg, Germany: Carl-Auer.

Schulz, K. F., & Grimes, D. A. (2002). Generation of allocation sequences in randomised trials: Chance, not choice. *The Lancet, 359,* 515–519. doi:10.1016/S0140-6736(02)07683-3

Sparrer, I. (2006). *Systemische Strukturaufstellungen. Theorie und Praxis* [Systemic structural constellations. Theory and practice]. Heidelberg, Germany: Carl-Auer.

Vermeersch, D. A., Whipple, J. L., Lambert, M. J., Hawkins, E. J., Burchfield, C. M., & Okiishi, J. C. (2004). Outcome questionnaire: Is it sensitive to changes in counseling center clients? *Journal of Counseling Psychology, 51,* 38–49. doi:10.1037/0022-0167.51.1.38

Received December 18, 2012
Revision received May 1, 2013
Accepted May 1, 2013 ■