UCSF

Title

Does treatment fidelity predict client outcomes in 12-Step Facilitation for stimulant abuse?

Permalink

Journal

ISSN

Authors

Guydish, Joseph
Campbell, Barbara K
Manuel, Jennifer K
et al.

Publication Date
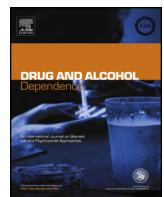
DOI

Peer reviewed

Does treatment fidelity predict client outcomes in 12-Step Facilitation for stimulant abuse?

Joseph Guydish [a,*], Barbara K. Campbell [b], Jennifer K. Manuel [c], Kevin L. Delucchi [d], Thao Le [a], K. Michelle Peavy [e], Dennis McCarty [b]

[a] University of California, San Francisco, Philip R. Lee Institute for Health Policy Studies, 3333 California Street, Suite 265, San Francisco, CA 94118, United States
[b] Oregon Health and Science University, Department of Public Health and Preventive Medicine, Portland, OR 97239, United States
[c] University of California, San Francisco, Department of Psychiatry, Building 20, Suite 2100, Room 2127, 1001 Potrero Avenue, San Francisco, CA 94110, United States
[d] University of California, San Francisco, Department of Psychiatry, Box 0984-TRC, 401 Parnassus Avenue, San Francisco, CA 94143, United States
[e] University of Washington, Alcohol and Drug Abuse Institute, Seattle, WA 98105, United States

ARTICLE INFO

ABSTRACT

Objective: This study examined the relationships between treatment fidelity and treatment outcomes in a community-based trial of a 12-Step Facilitation (TSF) intervention.

Method: In a prior multi-site randomized clinical trial, 234 participants in 10 outpatient drug treatment clinics were assigned to receive the Stimulant Abuser Groups to Engage in 12-Step (STAGE-12) intervention. A secondary analysis reviewed and coded all STAGE-12 sessions for fidelity to the protocol, using the Twelve Step Facilitation Adherence Competence Empathy Scale (TSF ACES). Linear mixed-effects models tested the relationship between three fidelity measures (adherence, competence, empathy) and six treatment outcomes (number of days of drug use and five Addiction Severity Index (ASI) composite scores) measured at 3 months post-baseline.

Results: Adherence, competence and empathy were robustly associated with improved employment status at follow up. Empathy was inversely associated with drug use, as was competence in a non-significant trend ($p = .06$). Testing individual ASI drug composite score items suggested that greater competence was associated with fewer days of drug use and, at the same time, with an increased sense of being troubled or bothered by drug use.

Conclusions: Greater competence and empathy in the delivery of a TSF intervention were associated with better drug use and employment outcomes, while adherence was associated with employment outcomes only. Higher therapist competence was associated with lower self-report drug use, and also associated with greater self-report concern about drug use. The nature of TSF intervention may promote high levels of concern about drug use even when actual use is low.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Multisite trials of behavioral interventions require treatment fidelity, or the delivery of treatment as intended, for internal and external validity. Fidelity monitoring assists standardization of treatment delivery across interventionists and sites, contributing to generalizability of results. This is particularly relevant for trials conducted in community settings where fidelity procedures support standardized treatment delivery across diverse settings and therapists (e.g., Ball et al., 2002). Treatment fidelity is also important in practice settings, where use of evidence-based treatments (EBTs)

is increasingly mandated or encouraged (Manuel et al., 2011), and where fidelity measurement is used to ensure that such practices are delivered (Miller et al., 2005).

Components of treatment delivery include therapist adherence and competence (Borrelli, 2011; Gearing et al., 2011). Adherence is the extent to which treatment sessions are delivered as stipulated by the intervention. Competence involves skillful delivery of treatment components, and the application of non-specific treatment skills such as empathy and alliance-building (Webb et al., 2010). Methods for monitoring fidelity include provider self-report and patient interviews (Borrelli et al., 2005), however, observation by independent raters is considered the "gold standard" (Baer et al., 2007; Gearing et al., 2011).

Comprehensive fidelity assessment examines the relationship of adherence and competence to patient outcomes, providing an

opportunity to identify a treatment's active ingredients (Borrelli, 2011). Studies of the relationship of adherence and competence to outcomes produce complex results (Webb et al., 2010). This may reflect methodological challenges, including the need for reliable and valid fidelity measures, the restricted range of adherence/competence found in clinical trials, the heterogeneity of analytic models used to study fidelity–outcome relationships (Barber et al., 2007), and small sample size and variability in patient symptom severity (Webb et al., 2012). Recommendations to address these challenges include statistical control for third variable confounds and analysis of interactions among adherence, competence and non-specific factors (e.g., therapeutic alliance; Barber et al., 2007; Webb et al., 2010).

In drug treatment, studies of fidelity–outcome relationships have reported mixed findings. Adherence to individual drug counseling (Barber et al., 2006) and adherence to multi-component therapy (Gibbons et al., 2010) were associated with better drug use outcomes only when therapeutic alliance was low, while adherence to cognitive–behavioral treatment predicted better drug use outcomes when therapeutic alliance was controlled (Hogue et al., 2008). Other studies found that greater adherence to motivational interviewing was associated with negative drug screens during treatment (Martino et al., 2008), that adherence to multi-dimensional family therapy predicted adolescent symptom reduction and improved family functioning (Hogue et al., 2006), and that adherence to brief strategic family therapy was associated with decreased adolescent drug use (Robbins et al., 2011).

Competence–outcome relationships have been weaker, and were not found for individual drug counseling (Barber et al., 2006), cognitive–behavioral therapy (Hogue et al., 2008), multidimensional family therapy (Hogue et al., 2008), or multi-component therapy (Gibbons et al., 2010). Poorer outcomes were associated with competent delivery of supportive–expressive therapy (Barber et al., 2008). One study found that motivational interviewing competence was positively associated with negative drug screens at 4 weeks (Martino et al., 2008). Overall, these studies demonstrated some support for the relationship of treatment fidelity, particularly adherence, to outcomes.

1.1. Twelve Step Facilitation (TSF)

Compatible with the 12-step philosophy that guides treatment in many programs, TSF is designed to promote 12-step engagement and participation (Donovan et al., 2011; Kaskutas et al., 2009; Project Match Research Group, 1997). It is considered an emerging EBT (Manuel et al., 2011), and has been used in both individual (Carroll et al., 1998a; Project Match Research Group, 1998) and group formation (Brown et al., 2002a; Kaskutas et al., 2009). Fidelity monitoring is important in TSF trials to ensure its distinction from similar, but untested 12-step oriented treatment in community practice.

This report examines the relationships of therapist adherence, competence and empathy to drug use outcomes in a community-based trial of TSF—the Stimulant Abuser Groups to Engage in 12-Step (STAGE-12; Daley et al., 2011; Donovan et al., 2013). Following completion of the trial, all treatment sessions were rated using a fidelity rating scale with demonstrated reliability and concurrent validity (Campbell et al., 2013). Ratings were used to evaluate the association of treatment adherence, competence and empathy to participant outcomes.

1.2. Overview of STAGE-12 study

Participants in the STAGE-12 study were adults seeking outpatient drug abuse treatment, who had used stimulants in the past 30 days and had a current diagnosis of stimulant abuse or dependence. In a two-group randomized design including 10 outpatient treatment programs, participants were randomly assigned within site to STAGE-12 (n = 234) or treatment as usual (TAU; n = 237).

The STAGE-12 intervention included 5 group and 3 individual sessions integrated into usual care. Group sessions were based on the Project MATCH TSF manual (Baker, 1998), adapted to a group format (Brown et al., 2002a, 2002b), and covered 12-step topics such as surrender and acceptance. Individual sessions, based on the introductory and termination sessions from the manual, reinforced participation in 12-step activities and incorporated an intensive 12-step referral procedure (Timko et al., 2006). Research assessments were at baseline (admission to treatment), 4 weeks, 8 weeks (end of treatment), and at 3 and 6 months post-randomization.

Study results, reported in Donovan et al. (2013), are summarized here. The follow-up rate for all participants was 70% at both 3 and 6 months. However at 3 months, the data point used in the present study, follow-up was lower in the STAGE-12 condition (65%) than in TAU (74%). Compared to TAU, STAGE-12 participants had a higher likelihood of abstinence from stimulant drugs at 4 weeks (OR = 3.34, 95% CI = 1.20, 9.28) and 8 weeks (OR = 2.44, 95% CI = 1.1, 5.86) but the difference was not significant at 3 months (OR = 1.78, 95% CI = 0.81, 3.90) or at 6 months (OR = 1.30, 95% CI = 0.60, 2.79). Compared to TAU, and including only participants with non-zero Addiction Severity Index (ASI) Drug Composite scores, STAGE-12 participants had significantly lower ASI Drug Composite scores at 3 months (t(206) = 2.02, p = .044). STAGE-12 participants also showed greater reduction in drug composite scores from baseline to 3 months (t(324) = −2.12, p = .035). Last, STAGE 12 participants had higher rates of engagement in 12-step programs throughout the 6 months of follow-up (Donovan et al., 2013).

2. Methods

2.1. Study design

To monitor fidelity of the intervention all STAGE-12 sessions (N = 999) were digitally recorded. For the present study, all digitally recorded files (DRFs) were rated for therapist adherence, competence and empathy. Some DRFs were corrupted, and one participant had no outcome data, leaving 965 usable recordings. Group sessions included multiple participants, so a single group DRF may result in ratings for more than one participant. Overall, there were 1124 usable session ratings for 215 STAGE-12 participants.

The present study concerned the association between the fidelity and outcome, so the analysis was constrained to cases having outcome data. Of 234 STAGE-12 participants, 153 (65%) completed the three-month follow-up interview (Donovan et al., 2013). The current paper reports findings for participants who completed the 3-month follow-up and had at least one usable DRF. Data reported below refer to 871 therapy sessions delivered to 151 participants.

2.2. Measures

2.2.1. Twelve Step Facilitation Adherence Competence Empathy Scale (TSF ACES). The TSF ACES (Campbell et al., 2013) is a fidelity rating scale based on prior fidelity measures (Carroll et al., 1998b, 2000) and on the STAGE-12 treatment manual. There are four content rating forms, one for group sessions (10 items) and three corresponding to STAGE-12 individual session 1 (10 items), session 2 (4–5 items) and session 3 (8–9 items). The number of items in sessions 2 and 3 vary based on whether the participant attended a 12-step meeting since the last session. The TSF ACES yields ratings for therapist adherence, competence, use of proscribed behaviors, empathy and global session rating. Adherence items assessed the degree to which the therapist delivered content prescribed for a specific session (e.g., "To what extent did the individual counselor encourage the participant to identify and agree to attend specific 12-step meetings?"). Corresponding competence items asked: "How well did the counselor handle this item?" Three counselor proscribed behaviors were rated: (a) material was overly structured; (b) used inappropriate self-disclosure; and (c) allowed focus to shift to irrelevant topics. The empathy item was adopted from the Motivational Interviewing Treatment Integrity scale (Moyers et al., 2005): "Overall, how well did the counselor understand or make an effort to grasp the clients' perspectives?" The global session rating was: "Overall, how well did the counselor conduct this session?" All items were rated on a 6-point Likert scale. Rating manuals and forms are found at: http://ctndisseminationlibrary.org/PDF/795_TSFACES.pdf.

Psychometric properties of the TSF ACES are discussed in Campbell et al. (2013). Inter-rater reliabilities, calculated using intraclass correlations, were .91 for mean adherence, .90 for mean competence, .83 for proscribed behaviors, .69 for global empathy, and .80 for global session rating. Cronbach's alpha was computed for multi-item adherence, competence, and proscribed behavior scales. As the number of adherence and competence items varied by type of session, a single coefficient was first computed for each session type, and these were averaged with weighting to account for number of sessions for each session type (i.e., individual sessions 1, 2, and 3). As proscribed behavior items were the same for all session types, no mean weighting was used. Alphas were 0.69 for adherence ratings and 0.71 for competence ratings, indicating acceptable internal consistency. The alpha for proscribed behaviors was poor (.47), suggesting that proscribed items measure different types of poor skill that do not co-occur. Content for individual session 2 differed depending on whether or not the participant attended a 12-step meeting since the last session. For 19 sessions where the participant did not attend a 12-step meeting, alphas were anomalous (adherence = −0.58, competence = 0.06), and these 19 sessions were excluded from analysis.

As the TSF-ACES measures are highly intercorrelated (Campbell et al., 2013), and to limit exposure to Type I error, we selected adherence, competence and empathy measures as predictors. Adherence and competence are multi-item measures and, in contrast to the single-item global session rating, may offer more stable estimates of underlying constructs. The empathy rating was retained to reflect a general counseling skill or common factor (Webb et al., 2010). The proscribed behavior measure was not used due to its lower internal consistency.

2.2.2. Treatment outcome measures. STAGE-12 research interviews, conducted at baseline and 3-month follow-up, included the Substance Use Calendar (SUC) and the ASI-Lite. The SUC measures days of drug use in the past month, with values ranging from 0 to 30, and has shown agreement with collateral informant interviews and urine assays (Ehrman and Robbons, 1994; Donovan et al., 2012). The ASI-Lite (McLellan et al., 1992) generates composite scores in seven areas (medical, employment, legal, alcohol, drug, social, psychological; McLellan et al., 1985). Composite scores measure problem severity during the 30 days preceding the interview, reflected as values from 0 to 1, and have been shown to be sensitive to treatment effects (McLellan et al., 1985). ASI composite scores are frequently non-normally distributed, with high proportions of 0 values for most composite scores and high proportion of 1 values for the Employment composite. Because measures with high proportions of 0 values are uninformative, and to limit exposure to Type I error, we included in this analysis only composites with less than 50% zero values at 3 months (Drug, Psychological, Medical, Family/Social, Employment). In this study the ASI drug composite core was correlated with the SUC measure of days of drug use ($r = 0.47$, $p < .0001$).

The ASI Drug Composite includes both "objective" (number of days of drug use in the past 30) and "subjective" items (Alterman et al., 2004). Subjective items are "how troubled or bothered have you been within the past 30 days by drug problems," and "how important is to you now is treatment for drug problems." Wertz et al. (1995) reported that decreases in drug composite scores were driven by decreases on the subjective items, while Alterman et al. (2004) found that both objective and subjective items decreased significantly over time.

2.2.3. Therapeutic alliance as a potential mediating variable. The Helping Alliance Questionnaire – Client Form includes 19 items concerning how the client perceived helpfulness and collaborative bonding with the therapist (HAq-II; Luborsky et al., 1996). Luborsky et al. (1996) reported internal consistency at 0.90, test–retest reliability of 0.78, and convergent validity with subscales of the California Psychotherapy Alliance Scale ranging from 0.38 to 0.71. As a measure of therapeutic alliance, we took the mean of the HAq-II items.

2.3. Procedures

Procedures for the STAGE-12 study are reported in Donovan et al. (2013). De-identified baseline and 3-month follow-up data were provided by the STAGE-12 team. Session DRFs included the participant research number so that session recordings and outcome data could be linked for the same participant. Institutional Review Boards at the University of California San Francisco and Oregon Health and Science University approved study procedures.

2.4. Fidelity rating procedures

Eight raters (excluding the master rater) rated the session DRFs. DRFs were randomly assigned to raters in sets of 20, and one DRF in each set was randomly selected for co-rating by the master rater to monitor inter-rater reliability. The master rater was a psychologist with both clinical and research experience, including experience developing fidelity rating scales, conducting fidelity monitoring, and training and supervising raters (e.g., Campbell et al., 2009). All raters met with the master rater in bi-weekly group conference calls to support rating consistency. Kappas of the co-rated sessions were monitored, and three consecutive co-rated kappas below .70 resulted in rater decertification and retraining. Weighted kappa coefficients for 59 co-rated sessions ranged from .31 to 1.00, with a mean of .69. Rater training and certification are discussed in Campbell et al. (2013).

2.5. Data analysis

Analyses assessed the relationships between the TSF ACES fidelity measures (adherence, competence, empathy) and drug treatment outcome measures (SUC and the five ASI composite scores), with the hypothesis that greater adherence, competence and empathy would be associated with better treatment outcomes. Therapeutic alliance, measured using the HAq II, was considered in the analysis as a potential mediator.

TSF ACES fidelity measures represent ratings of treatment sessions that occurred over 8 weeks of treatment post-randomization. The HAq-II was administered at the end of treatment, which occurred at 8 weeks post-randomization. ASI outcome measures were collected at the 3-month follow-up, or 12 weeks post-randomization. Consequently, both the predictors (TSF ACES scales) and the potential mediator (HAq-II measure of alliance) temporally preceded outcome measures (SUC, ASI composite scores).

Poisson regression modeling was used to test the relationship between each of the three fidelity measures and the SUC count data (days of drug use). Linear mixed-effects models were used to test the relationship between each of the three fidelity variables and the ASI composite scores (Drug use, Employment, Medical, Social, and Psychological). All models included the baseline value of each outcome, and age, gender, and race. Nesting of clients within site was controlled for as well. Nesting of clients within counselor was not controlled because clients received STAGE-12 group sessions from more than one counselor. As ASI composite scores were non-normally distributed, we converted them using a logit transformation.

We also tested whether therapeutic alliance (measured by the HAq-II) may be related to fidelity and outcome, using the same linear mixed-effects model approach. If therapeutic alliance was associated with both predictors and outcomes, it would be a potential mediator in the fidelity–outcome relationship. Last, as the primary analyses yielded inconsistent findings between the two drug use measures (SUC, ASI Drug Composite), we used the linear mixed-effects model to assess the relationship between fidelity and each of the four individual items that make up the drug composite score.

3. Results

3.1. Participant characteristics

Of 234 participants assigned to the STAGE-12 intervention, 151 completed the 3-month follow-up interview and had at least one usable STAGE-12 DRF. This sample had a mean age of 39.1 (SD = 10.4), 60% were women, and the racial/ethnic breakdown was 43% White, 40% African American, and 17% of other race. Persons of Hispanic ethnicity comprised 7% of the sample. Nearly half (49%) reported high school as the highest education achieved. At baseline, participants were employed full time (38%), part time (24%), unemployed (34%) and not seeking employment (5%). Participants included in the analyses ($n = 151$) did not differ from those excluded ($n = 83$) on these demographics. However, participants included in the analysis received more STAGE-12 sessions (mean = 5.8, SD = 2.0) than those not included (mean = 2.8, SD = 2.3, $t(232) = −10.12$, $p < .0001$) (see Table 1).

3.2. Relationship of fidelity measures to treatment outcome

Results are summarized in Table 2. Fidelity predictors are listed in the leftmost column, and treatment outcome variables are listed across the top. Reading down the first column shows that SUC number of days of drug use was inversely related to empathy (estimate = −0.36, $p = 0.025$). Participants who had sessions rated higher in empathy also had fewer days of drug use at 3 months. Competence ratings were also inversely related to days of drug use, however this did not reach statistical significance (estimate = −0.44, $p = 0.06$). The next column shows, unexpectedly, that the ASI drug composite was positively associated with adherence, competence, and empathy, such that participants receiving sessions rated higher on these characteristics also had higher ASI drug composite scores (greater problem severity) at 3 months. Employment was inversely related to fidelity measures, such that participants with sessions rated higher in adherence, competence and empathy had lower employment problem severity at 3 months.

Table 1
Baseline demographic characteristics for STAGE-12 study participants included ($n = 151$) and not included ($n = 83$) in fidelity–outcome analysis.

	Included in analysis $N = 151$	Not included $N = 83$	Total STAGE-12 sample $N = 234$
Age – mean (SD)	39.1 (10.4)	36.6 (9.3)	38.2 (10.0)
Gender			
Female	91 (60%)	54 (65%)	145 (62%)
Male	60 (40%)	29 (35%)	89 (38%)
Hispanic or Latino Ethnicity	11 (7%)	4 (5%)	15 (6%)
Race			
White	64 (43%)	42 (51%)	106 (45%)
African American	60 (40%)	27 (33%)	87 (37%)
Other[a]	26 (17%)	14 (17%)	40 (17%)
Highest education			
Less than high school	35 (23%)	17 (20%)	52 (22%)
High school graduate	74 (49%)	37 (45%)	111 (47%)
Some college	42 (28%)	29 (35%)	71 (30%)
Employment			
Full time	57 (38%)	26 (31%)	83 (35%)
Part time	36 (24%)	22 (27%)	58 (25%)
Unemployed	51 (34%)	30 (36%)	81 (35%)
Other[b]	7 (5%)	5 (6%)	12 (5%)
No. of STAGE-12 sessions, mean (SD)[c]	5.8 (2.0)	2.8 (2.3)	4.7 (2.5)

[a] Other includes: Alaskan = 3, Asian = 3, mixed = 19, native HI or Pacific Islander = 2, other = 13.
[b] Other employment categories were student (1), retired (7), and living in a controlled environment (4).
[c] 19 STAGE-12 participants received 0 sessions, and 0 values are included in the mean calculations.

Last, fidelity measures were not associated with Medical, Family/Social or Psychological ASI composite scores.

3.3. Relationship of fidelity measures to therapeutic alliance

The rightmost column of Table 2 shows results of models testing the relationship between the three fidelity measures and therapeutic alliance. Therapeutic Alliance was significantly associated with adherence (estimate = 0.41, $p = 0.002$) and competence (estimate = 0.37, $p = 0.002$), but not with empathy. However, therapeutic alliance was not significantly associated with any of the outcome variables (data not shown), so it was not included in the analysis as a potential mediator.

3.4. Relationship of fidelity measures to ASI drug composite items

The pattern of findings appeared contradictory. Greater empathy was associated with lower SUC days of drug use and with higher ASI drug problem severity. Greater competence was associated with lower SUC days of drug use (in a non-significant trend) and with higher ASI drug problem severity. Greater adherence, competence and empathy were each associated with lower employment problem severity at follow up. To explore why competence and empathy may be associated with fewer days of drug use and higher ASI drug composite scores we conducted further analysis of ASI drug composite data.

ASI drug composite scores are based on four items (McLellan et al., 1985). The first item asks number of days that each individual drug was used in the past 30. Summing days across drug categories and then subtracting the number of days the respondent used more than one drug offers a measure of number of days of any drug use. While it is less precise than the SUC, this offers an estimate of days of drug use based on the ASI, and the two measures of days of drug use were strongly correlated ($r = 0.72$, $p < .0001$). Three items complete the composite score: (1) How many days in the past 30 have you experienced drug problems (responses 0–30)? (2) How

Table 2
Parameter estimates, standard error (SE) and p values of models testing associations between STAGE-12 fidelity predictor, treatment outcome, and potential mediating variable ($N = 151$).[a]

Fidelity predictor variables	Treatment outcome variables						Potential mediating variable
	Days drug use past 30[b]	ASI drug composite[c]	ASI employment composite	ASI medical composite	ASI social composite	ASI psychological composite	Therapeutic alliance[d]
Adherence	−0.18 0.326 $p = 0.574$	1.05 0.378 $p = 0.006$	−0.77 0.384 $p = 0.048$	0.73 0.512 $p = 0.158$	0.04 0.245 $p = 0.860$	−0.13 0.233 $p = 0.585$	0.41 0.130 $p = 0.002$
Competence	−0.44 0.235 $p = 0.060$	0.75 0.335 $p = 0.027$	−1.15 0.305 $p < 0.001$	0.01 0.443 $p = 0.979$	0.03 0.203 $p = 0.866$	−0.10 0.193 $p = 0.596$	0.37 0.116 $p = 0.002$
Empathy	−0.36 0.160 $p = 0.025$	0.60 0.271 $p = 0.028$	−0.70 0.262 $p = 0.008$	−0.18 0.364 $p = 0.621$	0.20 0.166 $p = 0.231$	0.22 0.152 $p = 0.153$	0.07 0.094 $p = 0.433$

[a] Linear mixed-effects models were used to test relationships between fidelity ratings (adherence, competence, empathy) and ASI composite outcome variables and one potential mediating variable. Poisson regression models were used to test relationships between fidelity ratings and the SUC outcome variable. All models controlled for baseline value of the outcome, age, gender, and race, with clustering within site.
[b] Days of any drug or alcohol use as reported on the Substance Use Calendar.
[c] ASI is Addiction Severity Index.
[d] Therapeutic alliance as measured by the Helping Alliance Questionnaire (HAq II) at 8 weeks post randomization.

Table 3

Parameter estimates, standard error (SE) and p values of linear mixed-effects models testing associations between STAGE-12 fidelity predictors and items in the ASI drug composite score.[a]

Fidelity predictor variables	ASI drug composite score items			
	Number of days of drug use in past 30[b]	How many days in the past 30 have you experienced drug problems?	How troubled or bothered have you been within the past 30 days by drug problems?	How important to you now is treatment for drug problems?
Adherence	−1.66	2.43	0.82	0.95
	0.963	1.275	0.240	0.301
	$p = 0.088$	$p = 0.059$	$p < 0.001$	$p = 0.002$
Competence	−2.11	0.54	0.51	0.81
	0.783	1.109	0.213	0.256
	$p = 0.008$	$p = 0.628$	$p = 0.017$	$p = 0.002$
Empathy	−1.14	−0.31	0.30	0.60
	0.650	0.919	0.174	0.208
	$p = 0.081$	$p = 0.737$	$p = 0.090$	$p = 0.005$

[a] Linear mixed-effects models were used to test relationships between fidelity ratings (adherence, competence, empathy) and treatment outcome variables. All models controlled for baseline value of the outcome, and for age, gender, and race, with clustering within site.
[b] ASI number of days of drug use calculated by summing number of days reported for each individual drug, and then subtracting number of days reported using more than one drug.

troubled or bothered have you been within the past 30 days by drug problems (0 not at all to 4 extremely)? (3) How important to you now is treatment for drug problems (0 not at all to 4 extremely)? We disaggregated the drug composite score into its component items, and tested each item for its relationship to the fidelity measures.

Results in Table 3 show that the ASI measure for days of drug use (in the first column) was significantly and inversely associated with competence (estimate = −2.11, $p = 0.008$). We note that comparison of the first column in Table 2 with that in Table 3 suggests that both SUC days of drug use and ASI days of drug use were inversely, though not always significantly, related to all three fidelity measures. Days of drug use is the item that Alterman et al. (2004) identified as the "objective" measure within the drug composite score. The two "subjective" composite items, how troubled the participant has been by drug use and how important treatment for drug use is to them, are robustly and positively associated with adherence and competence, and the importance of treatment is associated with empathy. That is, while competence was inversely associated with days of drug use, it was positively associated with participant concern about their drug use.

4. Discussion

Employment is a common treatment outcome measure (Platt, 1995), and all three fidelity measures were significantly and inversely associated with ASI employment composite scores. That is, participants receiving more competent intervention also had better employment outcomes. Puzzling is that, given the association with employment, the three fidelity measures were (a) inconsistently associated with days of drug use and (b) inversely associated with ASI drug composite scores. Regarding the first point, a conservative view is that greater empathy (a non-specific treatment factor) was associated with fewer SUC days of drug use and that competence (a traditional fidelity measure) was associated with fewer ASI days of drug use. However, the significant competence–outcome relationship for ASI days of drug use is consistent with, and thereby lends strength to, the near-significant ($p = .06$) competence–outcome relationship for SUC days of drug use. Based on these observations, we conclude that both adherence and competence were associated with employment outcomes, and that competence was associated with drug use outcomes.

Regarding the point that fidelity measures were inversely associated with ASI drug composite scores, analysis of individual composite items revealed that persons receiving the intervention with greater competence reported fewer days of drug use while also reporting more concern about drug use. That this is artifact or measurement error seems unlikely, as days of drug use measured using the SUC also showed an inverse and, for empathy, a significant relationship. As far as we know, this is the first time that ASI objective and subjective items have shown opposite relationships to a predictor. This may be due to features of TSF and AA/NA programs that reinforce high subjective concern about drug use even in the absence of drug use, as well as the importance of treatment and including 12-step engagement. Maxims such as "one day at a time" keep the risk of drug or alcohol use at the fore even after years of sobriety. In any case, the positive relationship between therapist competence and ASI drug composite scores was driven by self-reported concern about drug use, and not by number of days of drug use.

Prior research has shown some support for adherence–outcome relationships (Hogue et al., 2008; Martino et al., 2008; Robbins et al., 2011) and weaker support for competence–outcome relationships (Barber et al., 2006; Gibbons et al., 2010, Hogue et al., 2008). The present study found a relatively stronger competence–outcome relationship and a relatively weaker adherence–outcome relationship. Webb et al. (2010) posit that the inconsistent relationships may be due partly to therapist responsiveness. Therapist adherence may vary in response to patient behavior (e.g., compliance, abstinence), and therapists may be more adherent with patients who fail to achieve abstinence or complete assignments. Our findings lend support to the assumption that fidelity measures of treatment delivery (competence) and non-specific treatment skill (empathy) improve some outcomes. Perhaps these components of treatment delivery reflect general therapist skills that effectively incorporate the responsiveness discussed by Webb et al. (2010).

Study findings suggest clinical and research considerations. As a clinical matter, funding and regulatory factors may determine whether the STAGE-12 intervention is applied in treatment settings. Our data suggest that, where it is applied, its effects can be strengthened by better therapist adherence, competence and empathy. Concerning future research on fidelity–outcome relationships, it may be helpful to include employment outcome measures, since findings for employment outcomes were most consistent in this study. Regarding 12-step research, further analysis of the objective and subjective ASI drug composite items is warranted. If the pattern observed here is borne out, then this measure may not be suitable for research or program evaluation of TSF interventions, and alternative measures of drug use or disaggregated

analysis of composite items (as done in this study) should be considered. Last, where fidelity data have been collected, they can be used to explore treatment outcomes. The usual analytic approach, when differences between study conditions are weak or null, is to conduct exposure analyses to assess whether persons receiving more of the experimental intervention have better outcomes. An analogous approach, taken here, is to conduct a fidelity analysis to assess whether persons receiving a better quality of intervention have better outcomes. While difficult in regular clinical practice, this may be done in clinical trials where fidelity monitoring procedures are routine but, apart from their role as intervention checks, do not inform analyses.

Limitations include the mixed results of the STAGE-12 study itself, the modest follow-up rate achieved in the STAGE-12 group at 3 months and, in the present analysis, measurement issues, sample size and reliance on self-report. The STAGE-12 intervention was superior to TAU on the primary outcome (days of stimulant use) during 8 weeks of treatment, but this difference was not significant at 3 months follow-up. If the intervention is considered ineffective on this basis, then fidelity of implementation appears unimportant. However, STAGE-12 participants also had lower 3-month ASI drug severity scores, a common treatment outcome measure, and greater engagement in 12-step activities over 6 months (Donovan et al., 2013), a factor previously associated with treatment outcome (Kaskutas et al., 2009). If ASI drug composite scores and 12-step engagement are considered important, then fidelity of STAGE-12 implementation also appears important. Last, data reported here show that, independent of outcomes as compared to TAU, better adherence and competence were associated with better outcomes within the STAGE 12 group.

The TSF ACES is a new measure, and psychometric properties would affect ability to detect fidelity–outcome relationships. The TSF ACES demonstrated good inter-rater reliability, good to excellent reliability for each of the summary measures, and good internal consistency for multi-item measures except the proscribed behaviors scale not used in this analysis (Campbell et al., 2013).

Intervention fidelity may be only modestly related to treatment outcome, and larger sample sizes may be helpful in identifying any consistent fidelity–outcome relationship. The sample size used here ($N = 151$) was larger than that in any of the 36 studies reviewed in Webb et al. (2010), where the median sample size was 33. Follow-up in the STAGE-12 condition was 65% at 3 months (Donovan et al., 2013), and Scott (2004) has shown that 30% loss to follow-up can result in significant and unpredictable bias. Further, those followed at 3 months received more STAGE-12 sessions than those not followed (5.8 vs. 2.8), and Kaskutas et al. (2009) reported a linear relationship between number of group TSF sessions attended and abstinence at follow-up. Consequently, findings reported here may be biased in unknown ways, and may not generalize to persons who leave treatment early or receive fewer treatment sessions. Dennhag et al. (2012) reported that, to determine counselor adherence and competence, analyses should have 5–10 sessions per patient and 4–14 patients per counselor. In our analyses there were 5.8 sessions per patient and a median of 5 patients per counselor. Last, outcome measures relied on patient self-report. It seems unlikely that self-report biases would account for findings since, in order to do so, participants receiving greater fidelity treatment would have to under-report days of drug use and at the same time over-report their level of concern about drug problems.

In this study higher competence and empathy in the delivery of a psychosocial intervention were associated with fewer days of drug use and better employment outcomes at 3 months post-baseline. Greater adherence to the protocol was associated with better employment outcomes only. Analysis of ASI drug composite items suggests that this measure may not be well-suited to TSF interventions or, possibly, that greater concern about drug use even

in the presence of lower levels of use is a marker for successful 12-step intervention. This study, reflecting one of the largest sample sizes for studies of the fidelity–outcome relationship (Webb et al., 2010) is suggestive, but not conclusive, that higher fidelity intervention is associated with improved treatment outcome.

Contributors

Dr. Guydish planned the work reported. He had full access to all of the data in the study and takes responsibility for the integrity and the accuracy of the data analysis. Dr. Campbell was responsible for rater training and supervision, for reporting psychometric properties of the TSF-ACES, and for writing sections of the Introduction and Discussion. Dr. Manuel contributed to the construction of reliability analysis, contributed to writing the Introduction and Methods, and served as a study rater. Dr. Delucchi provided statistical guidance throughout the course of the project, and Thao Le conducted analyses reported. Dr. Peavy was liaison to the STAGE-12 study team in the course of the work reported. Dr. McCarty provided scientific and administrative leadership for all work conducted at the Oregon health Sciences University.

Conflict of interest

No conflict declared.

References

Alterman, A.I., Cacciola, J.S., Koppenhaver, J.M., 2004. Addiction Severity Index composite scores: contribution of objective vs. subjective items to post-treatment change. J. Subst. Use 9, 214–223.

Baer, J.S., Ball, S.A., Campbell, B.K., Miele, G.M., Schoener, E.P., Tracy, K., 2007. Training and fidelity monitoring of behavioral interventions in multi-site addictions research: a review. Drug Alcohol Depend. 82, 107–118.

Baker, S., 1998. Twelve Step Facilitation for Drug Dependence. Yale University Psychotherapy Development Center, Substance Abuse Center, Training Series No. 3. Yale University, New Haven, CT.

Ball, S., Bachrach, K., DeCarlo, J., Farentinos, C., Keen, M., McSherry, T., Carroll, K., 2002. Characteristics, beliefs, and practices of community clinicians trained to provide manual-guided therapy for substance abusers. J. Subst. Abuse Treat. 23, 309–318.

Barber, J.P., Gallop, R., Crits-Christoph, P., Frank, A., Thase, M.E., Weiss, R.D., Gibbons, M.B., 2006. The role of therapist adherence, therapist competence, and the alliance in predicting outcome of individual drug counseling: results from the NIDA Collaborative Cocaine Treatment Study. Psychother. Res. 16, 229–240.

Barber, J.P., Gallop, R., Crits-Christoph, P., Barrett, M.S., Klostermann, S., McCarthy, K.S., Sharpless, B.A., 2008. The role of the alliance and techniques in predicting outcome of supportive–expressive dynamic therapy for cocaine dependence. Psychoanal. Psychol. 25, 461–482.

Barber, J.P., Sharpless, B.A., Klostermann, S., McCarthy, K.S., 2007. Assessing intervention competence and its relation to therapy outcome: a selected review derived from the outcome literature. Prof. Psychol. Res. Pract. 38, 493–500.

Borrelli, B., 2011. The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. J. Public Health Dent. 71, S52–S63.

Borrelli, B., Sepinwall, D., Ernst, D., Bellg, A.J., Czajkowski, S., Breger, R., DeFrancesco, C., Levesque, C., Sharp, D.L., Ogedegbe, G., Resnick, B., Orwig, D., 2005. A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. J. Consult. Clin. Psychol. 73, 852–860.

Brown, T.G., Seraganian, P., Tremblay, J., Annis, H., 2002a. Process and outcome changes with relapse prevention versus 12-Step aftercare programs for substance abusers. Addiction 97, 677–689.

Brown, T.G., Seraganian, P., Tremblay, J., Annis, H., 2002b. Matching substance abuse aftercare treatments to client characteristics. Addict. Behav. 27, 585–604.

Campbell, B.K., Fuller, B.E., Lee, E.S., Tillotson, C., Woelfel, T., Jenkins, L., Robinson, J., Booth, R.E., McCarty, D., 2009. Facilitating outpatient treatment entry following detoxification for injection drug use: a multi-site test of three interventions. Psychol. Addict. Behav. 23, 260–270.

Campbell, B.K., Manuel, J.K., Manser, S., Peavey, M., Stelmokis, J., McCarty, D., Guydish, J., 2013. Assessing fidelity of treatment delivery in group and individual 12-Step Facilitation. J. Subst. Abuse Treat. 44, 169–176.

Carroll, K.M., Nich, C., Ball, S.A., McCance, E., Rounsaville, B.J., 1998a. Treatment of cocaine and alcohol dependence with psychotherapy and disulfiram. Addiction 93, 713–727.

Carroll, K., Nich, C., Rounsaville, B.J., 1998b. Utility of therapist session checklists to monitor delivery of coping skills treatment for cocaine abusers. Psychother. Res. 8, 307–320.

Carroll, K.M., Nich, C., Sifry, R.L., Nuro, K.F., Frankforter, T.L., Ball, S.A., Fenton, L., Rounsaville, B.J., 2000. A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. Drug Alcohol Depend. 57, 225–238.

Daley, D., Baker, S., Donovan, D., Hodgkins, C., Perl, H., 2011. A combined group and individual 12-step facilitative intervention targeting stimulant abuse in the NIDA Clinical Trials Network: STAGE-12. J. Groups Addict. Recov. 6, 228–244.

Dennhag, I., Connolly Gibbons, M.B., Barber, J.P., Gallop, R., Crits-Christoph, P., 2012. How many treatment sessions and patients are needed to create a stable score of adherence and competence in the treatment of cocaine dependence? Psychother. Res. 22, 475–488.

Donovan, D.M., Daley, D.C., Brigham, G.S., Hodgkins, C.C., Perl, H.I., Floyd, A.S., 2011. How practice and science are balanced and blended in the NIDA Clinical Trials Network: the bidirectional process in the development of the STAGE-12 Protocol as an example. Am. J. Drug Alcohol Abuse 37, 408–416.

Donovan, D.M., Daley, D.C., Brigham, G.S., Hodgkins, C.C., Perl, H.I., Garrett, S., Doyle, S., Floyd, A.S., Knox, P.C., Botero, C., Kelly, T., Killeen, T., Hayes, C., Baumhofer, N.K., Seamans, C., Zamarelli, L., 2013. Stimulant abuser groups to engage in 12-Step (STAGE-12): a multisite trial in the NIDA Clinical Trials Network. J. Subst. Abuse Treat. 44, 103–114.

Donovan, D.M., Bigelow, G.E., Brigham, G.S., Carroll, K.M., Cohen, A.J., Gardin, J.G., Hamilton, J.A., Huestis, M.A., Hughes, J.R., Lindblad, R., Marlatt, G.A., Preston, K.L., Selzer, J.A., Somoza, E.C., Wakim, P.G., Wells, E.A., 2012. Primary outcome indices in illicit drug dependence treatment research: systematic approach to selection and measurement of drug use end-points in clinical trials. Addiction 107, 694–708.

Ehrman, R.N., Robbons, S.J., 1994. Reliability and validity of 6-month timeline reports of cocaine and heroin use in a methadone population. J. Consult. Clin Psychol. 64, 843–850.

Gearing, R.E., El-Bassel, N., Ghesquiere, A., Baldwin, S., Gilles, J., Ngeow, E., 2011. Major ingredients of fidelity: a review and scientific guide to improving quality of intervention research implementation. Clin. Psychol. Rev. 31, 79–88.

Gibbons, C.J., Nich, C., Steinberg, K., Roffman, R.A., Corvino, J., Babor, T.F., Carroll, C.M., 2010. Treatment process, alliance and outcome in brief versus extended treatments for marijuana dependence. Addiction 105, 1799–1808.

Hogue, A., Dauber, S., Samuolis, J., Liddle, H.A., 2006. Treatment techniques and outcomes in multidimensional family therapy for adolescent behavior problems. J. Fam. Psychol. 20, 535–543.

Hogue, A., Henderson, C.E., Dauber, S., Barajas, P.C., Fried, A., Liddle, H.A., 2008. Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. J. Consult. Clin. Psychol. 76, 544–555.

Kaskutas, L.A., Subbaraman, M.S., Witbrodt, J., Zemore, S.E., 2009. Effectiveness of making Alcoholics Anonymous easier: a group format 12-step facilitation approach. J. Subst. Abuse Treat. 37, 228–339.

Luborsky, L., Barber, J.P., Siqueland, L., Johnson, S., Najavits, L.M., Frank, A., Daley, D., 1996. The revised Helping Alliance questionnaire (HAq-II): psychometric properties. J. Psychother. Pract. Res. 5, 260–271.

Manuel, J.K., Hagedorn, H.J., Finney, J.W., 2011. Implementing evidence-based psychosocial treatment in specialty substance use disorder care. Psychol. Addict. Behav. 25, 225–237.

Martino, S., Ball, S.A., Nich, C., Frankforter, T.L., Carroll, K.M., 2008. Community program therapist adherence and competence in motivational enhancement therapy. Drug Alcohol Depend. 97, 37–48.

McLellan, A.T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grissom, G., Pettinati, H., Argeriou, M., 1992. The fifth edition of the Addiction Severity Index. J. Subst. Abuse Treat. 9, 199–213.

McLellan, A.T., Luborsky, L., Cacciola, J., Griffith, J., McGahan, P., O'Brien, C., 1985. Guide to the Addiction Severity Index: Background, Administration, and Field Testing Results. Publication No. ADM 85-1419. US Department of Health Human Services, Washington, DC.

Miller, W.R., Zweben, J., Johnson, W.R., 2005. Evidence-based treatment: why, what, where, when, and how? J. Subst. Abuse Treat. 29, 267–276.

Moyers, T.B., Martin, T., Manuel, J.K., Hendrickson, S.M.L., Miller, W.R., 2005. Assessing competence in motivational interviewing. J. Subst. Abuse Treat. 28, 19–26.

Platt, J., 1995. Vocational rehabilitation of drug users. Psychol. Bull. 117, 416–433.

Project MATCH Research Group, 1997. Matching alcoholism treatments to client heterogeneity: project MATCH posttreatment drinking outcomes. J. Stud. Alcohol Drugs 58, 7–29.

Project MATCH Research Group, 1998. Matching alcoholism treatments to client heterogeneity: project MATCH three-year drinking outcomes. Alcohol. Clin. Exp. Res. 22, 1300–1311.

Robbins, M.S., Feaster, D.J., Horigian, V.E., Puccinelli, M.J., Henderson, C., Szapocznik, J., 2011. Therapist adherence in brief strategic family therapy for adolescent drug abusers. J. Consult. Clin. Psychol. 79, 43–53.

Scott, C.A., 2004. A replicable model for achieving over 90% follow-up rates in longitudinal studies of substance abusers. Drug Alcohol Depend. 74 (1), 21–36.

Timko, C., DeBenedetti, A., Billow, R., 2006. Intensive referral to 12-step self-help groups and 6-month substance use disorder outcomes. Addiction 101, 678–688.

Webb, C.A., DeRubeis, R.J., Barber, J., 2010. Therapist adherence/competence and treatment outcome: a meta-analytic review. J. Consult. Clin. Psychol. 78, 200–211.

Webb, C.A., DeRubeis, R.J., Dimidjian, S., Hollon, S.D., Amsterdam, J.D., Shelton, R.C., 2012. Predictors of patient cognitive therapy skills and symptom change in two randomized clinical trials: the role of therapist adherence and the therapeutic alliance. J. Consult. Clin. Psychol. 80, 373–381.

Wertz, J.S., Cleaveland, B.L., Stephens, R.S., 1995. Problems in the application of the Addiction Severity index (ASI) in rural substance abuse populations. J. Subst. Abuse 7, 175–188.