Psychology ETDs

Electronic Theses and Dissertations

2-1-2016

Motivational Interviewing Treatment Integrity and Client Change: Using ROC Analysis to Explore the Relationship Between MI Fidelity Level and Drinking Outcome.

Daniel Fischer

Follow this and additional works at: https://digitalrepository.unm.edu/psy_etds

Daniel J. Fischer
Candidate

Department of Psychology
Graduate Unit (Department) Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Theresa B. Moyers, Ph.D., *Chairperson*

Sarah Feldstein Ewing, Ph.D.

Katie Witkiewitz, Ph.D.

Jennifer Hettema, Ph.D.

Motivational Interviewing Treatment Integrity and Client Change:

Using ROC Analysis to Explore the Relationship Between

MI Fidelity Level and Drinking Outcome.

by

Daniel J. Fischer

B.A., English, Seattle University, 2001

M.Ed., University of Notre Dame, 2003

Dissertation submitted in partial fulfillment

of the degree of Doctor of Philosophy

Psychology

University of New Mexico

Albuquerque, New Mexico

December 2015

DEDICATION

      Lea Anne, Jacob, and Violet, for offering support and inspiration in your own way and at your own time I will remain forever grateful. Of all that compels me forward, you are the greatest. I am a better person for sharing this journey with you.

ACKNOWLEDGEMENT

With heartfelt gratitude I would like to acknowledge Theresa Moyers, Ph.D., my dissertation committee chair and mentor. Thank you for your continued support and guidance throughout this process. Your example and your mentorship have taught me immeasurably about thinking and writing like a clinical scientist. Thank you for this opportunity. I am fortunate to have such a mentor and colleague.

To those who served on my dissertation committee: Sarah Feldstein Ewing, Ph.D., Jennifer Hettema, Ph.D., Katie Witkiewitz, Ph.D., and Kevin Vowels, Ph.D. Your questions and contributions have helped make this dissertation what it is. Thank you for your time, insight, and flexibility.

Finally, special thanks to Dan Matthews, Ph.D., for all that I have learned from you about *being* a clinical psychologist. Although not explicitly stated, you have influenced my thinking on this project as you have influenced much of what I have done and will do as a clinical psychologist.

Motivational Interviewing Treatment Integrity and Client Change:

Using ROC Analysis to Explore the Relationship Between

MI Fidelity Level and Drinking Outcome.

by

Daniel J. Fischer

B.A., English; M.Ed.

Ph.D., Psychology

Abstract

Those engaged in the research and practice of MI have shown interest in
treatment adherence as an indicator of effective MI and have expressed curiosity in the
threshold at which MI practice could be viewed as "good enough". The most widely
used and often cited of MI integrity measures are the Motivational Interviewing Skills
Code (MISC) and the Motivational Interviewing Treatment Integrity code (MITI). These
adherence tools share similar descriptive coding systems for therapist in-session
behavior. MI fidelity standards are often used as reference points for therapist
performance, yet practitioners rarely meet full criteria. Further, substandard ratings have
been associated with positive client change. These findings have elicited questions about
the necessary levels of therapist treatment adherence to promote client change and
suggested the need for empirically-derived fidelity standards. This study analyzed
existing data from a sample of Motivational Enhancement Therapy (MET) sessions from

Project MATCH (Matching Alcohol Treatments to Client Heterogeneity) that were audio

recorded and previously coded with the MISC. MI adherence variables were analyzed

along with client drinking outcomes to test the relationship between therapist fidelity and

client change. Therapist adherence was determined using behavioral codes common to

the MITI and MISC. Client change thresholds were determined using clinically

significant change standards developed by Jacobson and Truax. The relationships

between therapist adherence level and client change thresholds were examined using

Receiver Operating Characteristic (ROC) analysis. Findings showed mixed support for

the relationship between therapist adherence level and client drinking outcomes, but

yielded levels of therapist MI adherence associated with client changes in drinking

outcomes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

Background and history. The past several decades have seen an increased

emphasis in the identification and implementation of evidence based practices (EBPs)

within the field of clinical psychology. This movement towards a widespread

implementation of EBPs, the ultimate goal of which is to improve the effectiveness of

care, relies on empirical evidence from clinical research to establish best clinical

practices (Evidence Based Medicine Working Group, 1992). A subfield of research

within the larger EBP movement has focused on empirically supported treatments

(ESTs), clearly specified psychological treatments shown to be efficacious in controlled

research with a delineated population (Chambless & Hollon, 1998). Such research has

established a host of treatments for specified client disorders. For example, within the

field of addiction treatment, many ESTs have been identified to reduce substance use,

limit alcohol or drug related negative consequences, and increase client skills to support

ongoing recovery. Despite the existence of such treatments, challenges have been

encountered regarding effectiveness – that is, the ability of a treatment to produce the

desired effect within community-based clinical settings. Substandard treatment integrity

has been identified as one explanation for this gap between the efficacy of a treatment

and the effectiveness of its implementation. Researchers and clinicians have suggested

treatment adherence as one of the necessary conditions to ensure the highest probability

of client treatment success (Carroll & Rounsaville, 2007). Motivational Interviewing

(MI), an evidence-based and widely used intervention, serves as an illustration of the

challenges involved in establishing treatment adherence and integrity levels.

Motivational Interviewing – a brief, client-centered, and directive psychotherapy effective in facilitating client behavior change – is focused towards helping clients resolve ambivalence – the state of feeling two ways about something – in the direction of behavior change. Common examples of such behavior change include quitting cigarette use, improving one's diet or exercise, controlling drinking, or managing diabetes. MI has been shown to be an effective intervention for such behavior-oriented concerns. As a brief treatment, MI typically consists of one to four sessions and has been used as a stand-alone treatment, a pre-treatment booster to engage clients, or a complimentary addition to ongoing therapy (Miller & Rollnick, 2012). Theoretical assumptions of MI are client-centered, assuming the client to be the expert on his or her own experience and perspective. Within this stance, the clinician works with the client in an engaging and equal partnership, expressing an empathic and compassionate awareness and acceptance of the client's point of view. The clinician focuses the discussion on a targeted behavior, evokes from the client his or her own perspective towards change, and if appropriate assists in planning such a change (Miller & Rollnick, 2012). The therapist guides the client's exploration of change, providing support and encouragement and reinforcing the client's own reasons for change.

Expansion: effectiveness and efficacy. Since the first edition of *Motivational Interviewing* (Miller & Rollnick, 1991), research has grown to support MI as an effective intervention to target problem behaviors in a variety of settings. MI has been used broadly in the field of substance use treatment and has been found to be effective as a stand-alone treatment as well as in combination with other treatments (Dunn, Deroo, & Riviera, 2001; Burke, Arkowitz, & Menchola, 2003). It has been applied to changes such

as safer sex, diabetes, cardiovascular health, weight loss, and exercise (Hettema, Steele, & Miller, 2005). Beyond outpatient addiction counseling, MI has proven efficacious in a wide variety of settings including probation offices, emergency departments, primary care settings, schools, and community centers and churches (Miller & Rollnick, 2012).

Mechanisms of action. The latest research suggests that MI functions from a combination of relational components and technical components (Miller & Rose, 2009; Miller & Rollnick, 2012). The relational component includes principles that guide the therapist's way-of-being when working with a client. Often called "MI Spirit" this relational stance includes client-centered qualities such as acceptance, empathy, and collaboration in addition to other MI-specific qualities such as evocation, partnership, and compassion (Miller & Rose, 2009; Miller & Rollnick, 2012). The technical components of MI include specific therapist behaviors that are used to differentially attend to client statements in favor of changing (change talk; CT) rather than client statements in favor of the status quo (sustain talk; ST). Current perspectives from Miller and colleagues hold that basic counseling skills (such as reflections and open questions) as well as therapist behaviors that are MI-Adherent (MIA; such as emphasizing client control or reflecting client change talk) are associated with theoretically important client behaviors, namely client CT (Miller & Rose, 2009; Miller & Rollnick, 2012). Therapist MIA behaviors have been found to sequentially predict client change talk, and have been associated with increases in the frequency of client change talk and decreases in the frequency of sustain talk (Moyers & Martin, 2006). In-session rates of client CT have been found to predict improvements in client outcomes, specifically reductions in quantity and frequency of alcohol use (Bertholet, Faouzi, Gmel, Gaume, & Daeppen, 2010; Daeppen, Bertholet,

Gmel, & Gaume, 2007; Moyers, Martin, Houck, Christopher, & Tonigan, 2009).

Conversely, client in-session ST has been associated with both MI-Nonadherent (MINA)

behaviors from the clinician and decreases in client change at follow-up (Campbell,

Adamson, & Carter, 2010; Vader, Walters, Prabhu, Houck, & Field, 2010). Additionally,

specific training techniques, such as ongoing coaching and feedback on therapist MI

practice, have been found to predict increased rates of therapist MIA and decreased rates

of therapist MINA (Miller & Mount, 2001; Miller, Yahne, Moyers, Martinez, &

Pirritano, 2004; Moyers, Houck, Glynn, & Manuel, 2011). This research supports a

causal chain linking therapist training to in-session MIA behaviors, client in-session

change talk, and improvements in client outcomes at follow-up (Miller & Rose, 2009;

Moyers et al., 2009).

        Concerns about adherence and effectiveness. Concerns regarding MI adherence

first arose from concerning research findings. Therapist self-report of skill acquisition

post-training had only a modest relationship with observed therapist skill and no

relationship to client behaviors. Comparisons of training methods (workshop only,

workshop plus coaching, workshop plus coaching and feedback) found that therapist self-

report of MI skill was unrelated to demonstrated in-session behaviors and those who

received workshop training alone – the standard training practice – showed the least

improvement in skill at 12-months follow up (Miller et al., 2004).

        Scholars expressed concerns towards an over-emphasis on treatment adherence

when no data-based adherence measure existed and therapist self-report remained a

barrier to the establishment of meaningful in-session treatment integrity measurement

(Miller, 2001; Rollnick, 2001; Shmeige, Broaddus, Levin, & Bryan, 2009). Meta

analyses of MI outcome studies found that adherence monitoring was often not conducted

or not reported (Burke et al., 2003; Dunn et al., 2002; Hettema et al., 2005). One review

found that 71% of MI outcome studies failed to report treatment adherence measures, and

another review came to the conclusion that studies without fidelity reporting should be

excluded from reviews of MI outcome studies (Dunn et al., 2001; Hettema et al., 2005).

Misconceptions regarding MI brought added attention to the gap between

therapist self-report and demonstrated skill, further emphasizing the need for treatment

integrity monitoring (Miller & Rollnick, 2009). Many practitioners reported that MI was

interchangeable with what they already were practicing, whether it was client-centered

therapy or cognitive therapy; many believed MI was no different from the transtheoretical

model, or the decisional balance (Miller & Rollnick, 2009). Such discord in both

research and practice communities underlined the need for valid and reliable integrity

measures by which therapist MI performance could be judged.

MI adherence measures: MISC & MITI. In response to the need for MI integrity

measures, several coding systems were developed to evaluate therapist and client

interactions and provide feedback regarding MI (Miller & Mount, 2001; Miller et al.,

2004). These coding systems have approached integrity rating through two basic

processes: identifying specific in-session behaviors and applying global ratings across

sessions.

The Motivational Interviewing Skills Code (MISC) was the first coding system to

identify therapist and client behaviors and provide specific information on therapist

fidelity to MI principles and practices (Miller, 2000; Moyers, Martin, Catley, Harris,

Ahluwalia, 2003). The MISC exclusively and exhaustively labeled therapist and client

behaviors; all language within an MI session was included. MISC codes described various MI-consistent, MI-inconsistent, and neutral therapist behaviors as well as different responding styles of the client: change talk, sustain talk, or neutral language (Miller, 2000). Some versions of the MISC have also included several global measures of therapist behaviors: empathy, acceptance, "MI Spirit" (an average of evocation, collaboration, and autonomy support), and direction (Houck, Moyers, Miller, Glynn, & Hallgren, 2010). The MISC has been applied extensively in MI mechanisms research, and was used to establish sequential probabilities between therapist MICO behaviors and client CT responses (Moyers et al., 2009). Although the MISC provided a rich sample of in-session behavior data, it was found to be extremely labor intensive, requiring extensive training, multiple coding passes, and several hours of coding per therapy session.

The Motivational Interviewing Treatment Integrity code (MITI) was designed as an economical clinician-focused system for judging therapist MI treatment fidelity. An abbreviated version of the MISC, the MITI coded therapist behavior within a randomly selected 20-minutes segment of therapy. It was developed using factor analysis to extract elements of the MISC. This resulted in a coding system that included both descriptive behavior monitoring and gestalt global characterization. The broad behavior domains included: MI-adherent (MIA) behaviors, MI-Nonadherent (MINA) behaviors, giving information, questions, and reflections (Moyers et al., 2005). Behaviors such as giving information, open and closed questions, and simple and complex reflections captured general counseling skills central to effective MI practice. MI-adherent behaviors such as emphasizing client control or choice, client affirmation, or seeking collaboration with the client identified therapist behaviors characteristic of effective MI practice. MI-

Nonadherent behaviors characterized therapist behaviors contradictory to an MI approach. These behaviors included, confronting, directing, warning, or persuading the client. From these descriptive categories, summary statistics characterized patterns in therapist behavior across sessions. The summary statistics included the percentage of open questions (%OQ; total open questions/total questions – both open and closed questions), percentage complex reflections (%CR; total complex questions/total reflections – both complex and simple reflections), reflection to question ratio (R:Q; total reflections/total questions), and percent MIA (%MIA; total MIA behaviors/total MIA and MINA behaviors). The MITI also provided five-point Likert-type global ratings of therapist behaviors such as evocation, partnership, empathy, autonomy support, and direction. As with the MISC, evocation, partnership, and autonomy support global measures were collapsed together to produce an "MI Spirit" global measure.

MITI-derived therapist behavior ratios have been used to evaluate therapist performance against integrity standards (Table 1). These integrity standards were expert-based thresholds, providing benchmark criteria for both initial proficiency as well as ongoing competency. Such guidelines have served as useful standards to judge MI-performance, yet they were not empirically derived and represented what an expert in MI might anticipate to observe in a session with a high level of MI treatment adherence. The expert-based proficiency and competency thresholds represented aspirational rather than empirical standards.

Unanswered Questions. Underlying the issue of MI treatment adherence is the question: *When is MI good enough?* A logical and thoughtful response to this question might be: *Good enough for what* (Miller & Rollnick, 2012)*?* Different goals may call for

different levels of fidelity. One might want to measure changes in MI skill before and after training, or whether a practitioner can demonstrate proficiency after sustained practice, coaching, and feedback. One might be interested to know if a particular therapist can practice MI at a level associated with client post-treatment change. In clinical settings higher ratings of treatment adherence should be positively correlated with client outcomes such that as integrity increases, so does client treatment outcome. It would be important to know if there was a plateau in treatment integrity, beyond which further increases in fidelity were met with diminishing gains in client outcome. In treatment contexts where ongoing training, coaching, and supervision represent a commitment of resources on the part of an organization, the point at which improvements in treatment adherence no longer lead to increased improvements in client outcome would be important to know.

MITI and MISC ratings of therapists have often failed to meet the benchmark criteria set for beginner proficiency and competency, whether targeting health behavior change or substance use (Bohman, Forsberg, Ghaderi, & Rasmussen 2013, Carels et al., 2007; Forsberg, Ernst, & Farbring, 2011; Cimini et al., 2009; D'Amico et al., 2012). Therapist global ratings and behavioral measures of open-questions or total reflections have been closer to existing benchmark standards than therapist behavior summaries, a finding common across substance abuse counselors, medical residents, nurses, or peer-interventionists (Lindqvist et al., 2013; Smith et al., 2012; Thush et al., 2009; Tollision et al., 2008). Improved global scores or increased basic counseling skills, like open questions or reflections, have typically been easier to achieve than improvement in the use of MI-specific counseling skills, such as reduced MINA behaviors, increased MIA

behaviors, or the use of more complex reflections. Therapist adherence ratings have typically improved when individuals received ongoing coaching and feedback on their MI practice (Daeppen et al., 2007; Carpenter et al., 2012, Miller & Mount, 2001).

The findings from MI adherence literature have been mixed regarding the relationship between treatment adherence and client outcomes. Sub-standard adherence scores have been associated with positive client treatment outcomes. This general finding is observed across interventions focused on various behavior change targets (Ang et al., 2012; Carels et al., 2007; Hansen et al., 2012; McCambridge, Day, Thomas, & Strange, 2011). Although higher therapist MITI scores generally shared a positive association with client outcomes, the average scores were often below benchmark standards.

No clear explanation has been given for the observed differences in MI treatment adherence ratings, nor has existing research explored the relationship that these ratings have to client outcomes. Relatively few studies have reported MI fidelity scores *as well as* client behavioral outcomes (Turris et al., 2009; Smith et al., 2012; Forsberg et al., 2011, Evangeli, Longley, Swarts, 2011). Training studies typically report only therapist fidelity levels while treatment outcome studies focus primarily on client outcomes, mentioning only that fidelity levels were met (Burke, Arkowitz, & Menchola, 2003; Hettema, Steele, & Miller, 2005; Madson, Loignon, & Lane, 2009). For those studies that did provide both treatment adherence and outcome data, global scores were often at or near the proficiency and competency standards, but behavior summary scores are far below existing standards (Turris et al., 2009; Smith et al., 2012; Forsberg et al., 2011, Evangeli, Longley, & Swarts, 2011). The extent to which therapist performance-based

integrity scores are related to client outcomes and at what level, if any, treatment fidelity

can identify subsequent client behavior change has remained unknown.

Receiver Operating Characteristic (ROC) analysis provides a methodology that

allows for the preservation of individual therapists and corresponding client performance

as well as the identification of optimal fidelity levels for classifying client change. Based

on signal detection theory and designed by radar engineers in WWII for purposes of

enemy threat detection, ROC analysis has been used in psychology to study stimuli

detection and in other health related fields to test the utility of decision making

instruments (Hajian-Tilaki, 2013). Also called relative operating characteristics, ROC

analysis is a graphical plot that illustrates the utility of binary classification systems by

testing the true positive rate and the false positive rate of signal detection as a

discrimination threshold varies. Although ROC analysis is commonly used as a decision-

making tool and has long been used in psychological research, this methodology has not

been applied as a decision making tool for identifying effective therapy via therapist

treatment fidelity level performance on identifying client change. It is a methodology

well-suited to exploring the relative performance of treatment integrity levels on

identifying client treatment outcome.

Aims. The aims of this study are: (1) to determine if MI treatment adherence

standards are empirically related to clinically significant client change, and (2) to employ

ROC curve analysis in a novel context. By exploring the utility of ROC curve analysis in

the context of therapist MI treatment adherence and client drinking outcomes, this project

aims to identify levels of MI adherence that can serve as indicators of treatment

effectiveness.

Methods

The current study was a secondary analysis exploring the relationship between therapist MI adherence and client outcomes for the treatment of alcohol use disorders. It tested the novel application of a commonly used signal detection methodology to assess the performance of therapist integrity measures on identifying clinically significant change status for clients engaged in Motivational Enhancement Therapy for Alcohol Use Disorders.

Research Plan. The parent study, which explored the relationship between therapist and client in-session behavior and the causal influence of these behaviors on subsequent client drinking outcomes, was a secondary analysis of Project MATCH MET data (Moyers et al., 2009).   First sessions of MET were coded with the MISC 2.0 to produce a dataset of descriptive codes for therapist and client therapeutic interactions. Sequential analyses of the data established a causal link between therapist MI-consistent behaviors (MICO) and client change talk (CT). Multilevel analyses found that therapist MICO as well as client CT levels predicted changes in client drinking levels. Project MATCH (Matching Alcohol Treatment to Client Heterogeneity), was a multi-site, randomized, controlled, clinical trial testing multiple hypotheses regarding benefits of matching specific client characteristics to specific alcohol treatments. In Project MATCH, clients were randomly assigned to one of three different ESTs for alcohol use disorder: Motivational Enhancement Therapy (MET; a variant of MI), Cognitive Behavioral Coping Skills Therapy (CBT), or Twelve Step Facilitation (TSF). MET and CBT/TSF included four or 12 sessions, respectively, over a 12-week period. Participant drinking outcomes were followed for one year post-treatment. The main outcomes from

the study showed that all client groups reduced alcohol use and that there was little difference in client outcomes across different treatment approaches (Project MATCH Research Group, 1998).

Project MATCH used a four-session MI intervention with study participants. In the first session, client concerns regarding the target behavior were discussed, and if the client expressed readiness to change, a change plan was established. In subsequent sessions, the target behavior was discussed along with feedback from earlier assessments. In these MI sessions, the therapist worked collaboratively with the client, eliciting the client's perspective towards change, supporting the client's autonomy in choosing whether or not to change, and worked towards resolving client ambivalence in the direction of decreasing client alcohol use.

In the current project, therapist integrity measures were compared with client drinking outcome data for the sampled sessions from Project MATCH. Client outcome data included percent days abstinent (PDA), drinks per drinking day, (DDD) and negative consequences of drinking. Outcome variables of PDA and DDD were gathered through the Form-90 (Tonigan, Miller, and Brown, 1997) and negative consequences of drinking were gathered through the Drinker's Inventory of Consequences (DrInC; Miller, Tonigan, and Longabaugh, 1995). Client alcohol use data was collected at various time points, and data from baseline, week-one, week-two, month-three, and month-six of study involvement was used to in the current study analyses. A total of 118 client first sessions were included for analysis. This sample was limited to audible recordings of sessions from community-based research settings that received permission from their local IRB to share data with another institution.

Participants. The current project used a sample of the total participants from the

Project MATCH MET arm. Of the 118 participants that appear in these sessions, 91

(77%) were male, with 60% assigned to aftercare and 40% treated in an outpatient

setting. The therapists that conducted the MET sessions consisted of masters and

doctoral level practitioners who were trained in MET and demonstrated pre-treatment

training benchmarks before conducting MET with study participants. Throughout the

Project MATCH these therapists received ongoing coaching and support and were

monitored to insure adequate treatment fidelity.

Analysis Plan. Receiver operating characteristic analysis, also called ROC curve

analysis, has been widely used as a methodology for organizing classifier variables and

illustrating their performance in decision-making tasks. Typically, the classifiers are not

binary and thresholds are necessary to dichotomize continuous variables. With two

binary classifiers, there are four possible outcomes: true positive, false positive, true

negative, and false negative. Based on signal detection theory, this classification system

involves the sensitivity and specificity of predictor variables, plotting true positive rates

(TPR; or sensitivity) and false positive rates (FPR; 1 – specificity) along the *y*-axis and *x*-

axis, respectively. A diagonal line running from the lower left to the upper right signifies

random performance, an equal probability of either true or false positive results. Any

point plotted above the lower left to upper right diagonal signified better than chance

classification. One plot point on the ROC graph is viewed as better than another if it is

further to the NW of the graph, suggesting a larger TPR and smaller FPR. The ROC

curve is the line formed by connecting discrete plotted points, representing the tradeoff of

FPR and TPR. By plotting TPR and FPR rates in two-dimensional space, ROC analysis

allows for the evaluation of a variable's performance as a classifier under various

conditions. Within the current project, TPR was the rate of correctly identifying a

positive treatment outcome based on therapist fidelity rate and FPR was the rate of

incorrectly identifying a positive treatment outcome based on therapist fidelity level.

 A unique ROC curve was generated to test the performance of each fidelity

measure on identifying client change status in an outcome measure at a particular time

point. ROC curves were represented in two-dimensional space, illustrating the

performance of the fidelity measures in terms of true positive rate and false positive rate.

A variables' performance was judged to be better or worse depending on its area under

the curve (AUC). The AUC can be understood as the overall performance of the

identifier variable such that the AUC represents the probability at any point along the

ROC curve of correctly distinguishing between any randomly selected pair of scores, one

from the population of positive cases, the other from the population of negative cases.

Once the ROC analysis was run, each observed score along the ROC curve was ranked

according to differences between TPR and FPR scores. This index of scores has been

called the Youden's Index and has been used to identify the point along the ROC curve

with optimal performance (Hajian-Tilaki, 2013). The AUC and the optimal cutoff along

the Youden's Index provide the most useful measures of the overall performance of the

identifier variable and the threshold at which the variable has optimal performance.

 ROC curve analysis required that continuous client outcome variables such as

change in the quantity of alcohol consumed, the frequency of alcohol consumption, or the

number of alcohol related negative consequences be dichotomized into change or no

change status. Jacobson and Truax's (1991) clinically significant change (CSC) criteria

were used to establish dichotomous change thresholds for each outcome variable. This

methodology has been used broadly in behavior change literature and has shown to be a

robust and conservative method for providing a change index that is both clinically

meaningful as well as reliable (McGlinchey et al., 2002; Atkins et al., 2005). More

specifically, Jacobson and Truax's approach has been used previously in alcohol and

other drug treatment research (Cisler et al., 2005; Maisto et al., 1996). Clinically

significant change is defined as post-treatment client functioning that meet the following

criteria: (a) existing outside of the range of the dysfunctional population (that is, being at

least two standard deviations away from the mean of the dysfunctional population), (b)

existing within the range of normal functioning (again, being within two standard

deviations of the mean of the normal population), or (c) existing closer to the mean for

the functional population than the dysfunctional population (Jacobson & Truax, 1991).

Jacobson and Truax developed a reliable change index (RC) that tests the statistical

likelihood of a clinically significant change occurring by chance. This RC score divides

the difference between pre-treatment and post-treatment scores by the standard error of

the difference between the two scores. An RC score that is larger than 1.96 is statistically

unlikely to have occurred by chance, and is therefore viewed as reliable change (Jacobson

& Truax, 1991).

    Power. Reviews of motivational interviewing found that treatment effect sizes

varied by study and target behavior. As an intervention for problematic drinking, the

effect size for MI has varied, between a low of $d = 0.26$ (Hettema et al., 2005) to a high

of $d = 0.60$ (Vasilaki et al., 2006). Several studies have found an effect size ranging

between $d = 0.41$ (Hettema et al., 2005) and $d = 0.47$ (Burke, et al., 2003), which can be classified as a moderately large magnitude.

Determining the necessary sample size to ensure adequate power for ROC analyses required the consideration of several factors. First, it depended on whether the intent was to establish that a ROC curve was different from chance or different from another ROC curve. Secondly, it depended on the base rate for the intended outcome in the general population. Previous studies found that determining whether an ROC curve was statistically different from chance required a smaller sample size than determining if one ROC curve was different from another (Bradley & Longstaff, 2004). Additionally, the further a base rate or class probability was from 50%, the greater the sample size required to prove an effect was statistically different from chance (Bradley & Longstaff, 2004). Previous analysis of Project MATCH data using clinically significant change criterion found that at three months, 51% of participants showed change in PDA, 47% showed change in negative consequences (DrInC scores), and roughly 25% showed change in DDD. Such rates fell at one-year follow up, with 33% showing change in PDA, 35% showing change in negative consequences, and 15% showing change in DDD (Cisler et al., 2005). Mindful of these change rates, data from the 118 available therapy sessions allowed adequate power to test whether or not a resulting ROC curve was different from chance, but was underpowered to test if one ROC curve was statistically different from another (Bradley & Longstaff, 2004).

Results

Descriptive statistics. Descriptive statistics for client drinking outcomes are shown in Table 2. The percentage of clients who met criteria for clinically significant

change is reported in Table 3. Descriptive statistics for therapist MI fidelity variables are

shown in Table 4. Of note, MI integrity ratings were much lower than the expert-

recommended proficiency and competency ratings (Tables 1 and 4). The one exception

to low average treatment integrity was the R:Q ratio, which met beginner proficiency

standards. Of the sample of coded sessions, no single session met all criteria for MI

proficiency or competency, although several met multiple in-session summary score

standards.

   ROC curve analysis using therapist adherence variables to identify clinically

significant change in drinking outcomes. Primary analyses tested the utility of therapist

MI fidelity for identifying CSC in various client drinking outcomes. Amongst the full

sample of clients, several MI adherence measures demonstrated significant performance

in identifying CSC in client drinking outcomes. Three of these variables identified CSC

in DDD at two-weeks of treatment. The variable with greatest performance was %OQ

identifying client CSC in DDD at two-weeks of treatment (Figure 1, Table 5). The

optimal level of %OQ in identifying CSC in DDD at two-weeks of treatment was %OQ =

30.82% (Figure 1, Table 11). The second of these variables was %MIA, significantly

identifying CSC in DDD at two-weeks of treatment (Table 5). The optimal performance

level of %MIA for identifying two-week changes in DDD was %MIA = 77.04 % (Table

11). The third of these variables identifying CSC in DDD at two-weeks of treatment was

the R:Q (Figure 1; Table 5). This variable showed optimal performance in identifying

CSC in DDD at two-weeks of treatment when R:Q = .927 (Table 11). The %OQ variable

also performed well in identifying CSC in DDD at six-months of treatment (Figure 2,

Table 5). Similar to the two-week DDD time point, %OQ showed optimal performance in identifying CSC in DDD at six-months of treatment when %OQ = 30.82% (Table 11).

ROC curve analysis using therapist adherence variables to identify reliable change in drinking outcomes. Follow up analyses were run to test the utility of therapist MI fidelity measures on identifying client change at thresholds less stringent than the conservative clinically significant change thresholds. Reliable client change (RC), described above, was used as an alternative threshold for client change. Among the full sample of clients, %OQ performed well in identifying RC in DDD at six-months of treatment (Figure 3, Table 5) with an optimal performance level of %OQ = 30.62% (Table 11). The %MIA performed well in identifying RC in PDA at various time points. At two-weeks of treatment %MIA identified RC in PDA (Figure 4, Table 9) with an optimal performance level of %MIA = 71.01% (Table 11). At three-months of treatment, %MIA identified RC in PDA (Figure 5, Table 9) with an optimal performance level of %MIA = 70.29% (Table 11).

ROC curve analysis using client change language to identify changes in drinking outcomes. Additional secondary analyses were run to test the utility of ROC curve analysis and the performance of client change language in identifying subsequent client changes in drinking outcomes. The percentage of client change talk (%CT; Total Change Talk / the sum of Total Change Talk and Total Sustain Talk) was used as a measure of in-session client change language. Amongst the full sample of clients, %CT performed well in identifying CSC change in alcohol-related consequence at multiple time points. At three-months of treatment, %CT performed well in identifying CSC changes in alcohol-related consequences (Figure 6, Table 12) with an optimal performance level of %CT =

.7869 (Table 13). These findings were replicated at six-months of treatment, with %CT identifying CSC in alcohol-related consequences (Table 12) and an optimal performance level of %CT = .7869 (Table 13).

Using the less stringent, but still statistically significant change threshold of reliable change, client change talk percentages identified reliable change in alcohol-related consequences at both three and six-months of treatment. Among the full sample of clients, %CT performed well in identifying RC in alcohol-related consequences at three-months of treatment (Figure 7, Table 12) with an optimal performance level of %CT = 78.69% (Table 13). Similar results for %CT identifying RC in alcohol-related consequences were found with the full sample of clients at six-months of treatment (Table 12) with an optimal performance level of %CT = 72.00% (Table 13).

Discussion

Considering the broad application of MI to different contexts and the attention given to MI training and effective implementation, this study provides the answer to an ongoing question within the clinical and research communities: when is MI good enough (Miller & Rollnick, 2012)? The goal of this study was to establish evidence-based therapist performance standards that were related to meaningful client change. Therapist adherence levels for in-session behaviors such as %MIA, R:Q, and %OQ, as well as client levels of %CT, identified meaningful client changes in DDD, PDA, and alcohol-related consequences at various time points. A repeated finding throughout this study was that therapist integrity levels associated with client change were often much lower than the existing expert-derived standards for MI adherence currently in use. This suggests that current expert-derived thresholds for MI fidelity are overly stringent and

that clients might do just as well with lower levels of critical MI behaviors on the part of

the therapist.

   Treatment integrity. Clinicians from Project MATCH, a population of highly

skilled, well trained, and closely supervised therapists, showed average MI treatment

integrity scores that were below the expected thresholds for proficiency and competency.

Despite these lower than expected fidelity levels, treatment was effective, with high rates

of change in the quantity of alcohol consumed, the frequency of alcohol consumed, and

the amount of negative alcohol-related consequences. These changes were observed

early in treatment and were sustained throughout treatment and follow up. Overall, this

suggests a level of therapist fidelity that, although below expert standards, provides an

effective treatment context sufficient for clients who are contemplating change.

   The adherence variables of %MIA and %OQ performed best in identifying

changes in client drinking outcomes. Because these variables are summaries of different

therapeutic micro-skills, it is unclear if the observed performance was due to more of one

behavior (MIA or OQ), less of another behavior (MINA or CQ), the relative comparison

of the two, or perhaps the artful use of theoretically meaningful behaviors within a

therapeutic context. For both %MIA and %OQ, the range of optimal performance fell

below the existing competency and proficiency standards. These lower performance

thresholds suggested that behaviors often eschewed within an MI context may have a

benign therapeutic effect in small doses when bolstered by larger amounts of MI-

consistent behaviors. This observation is consistent with earlier findings that therapist

MINA behaviors have, paradoxically, been found to enhance client engagement in

therapy sessions (Moyers, Miller, and Hendrickson, 2005).

The ratio of R:Q showed modest performance in identifying client change and was the only adherence variable that demonstrated an optimal performance at the level of existing proficiency standards and underlined the importance of balancing the use of reflections with questions in effective MI. Therapist %CR did not perform well in identifying subsequent client change. Such null findings may have been due, in part, to relatively lower reliability on this particular fidelity measure (Moyers et al., 2009).

Taking these findings into consideration, a new set of fidelity standards can be created based on study outcomes. These standards, which are empirically derived represent performance thresholds that fall below the existing competency and proficiency standards but were associated with subsequent client change. Although a causal relationship cannot be ascribed to such findings, they show promise towards the identification of fidelity standards associated with successful MI implementation. These sufficiency standards can be characterized as MI that has a %CR of at least 30%, %OQ of 30% or greater, a R:Q which is at least 0.9, and a %MIA that is 75% or greater.

Recent MI integrity publications by experts in the field are relevant to results from this study. In Miller and Rollnick's third edition of the *Motivational Interviewing* (2012), the authors discussed performance thresholds. They stated that training and supervision should be approached through a criterion-based methodology rather than simply an experience-based practice. The authors acknowledged that MI-adherence level may vary from one situation to another, with some contexts requiring much higher levels of integrity than others. Ultimately, this issue was framed as an empirical question that could be answered through research. These ideas are consistent with finding from the current study, which suggested that a relatively lower level of MI integrity than suggested

by experts performed well in the prediction of subsequent client change amongst clients

who self-referred for alcohol treatment.

Miller and Rollnick have also recently published on the role that treatment

integrity plays in the effectiveness or ineffectiveness of behavioral interventions (Miller

& Rollnick, 2014). In this publication, they compared treatment integrity to the dose

effect of a vaccine or an antibiotic; with weak doses often leading to ineffective

treatment. To carry this metaphor further, too much of a dose can sometimes lead to

harmful side-effects. The challenge is to apply the correct dose, monitor client responses,

and adjust the dose as needed. This approach to treatment integrity was also consistent

with the current study, which suggested a "dose" or fidelity level sufficient to identify

subsequent client change within the current context of self-referred clients seeking to

change their alcohol use. It remains unknown what appropriate levels of fidelity are

necessary to identify subsequent client change in other treatment contexts and for other

targeted behaviors.

Client characteristics. This study also provided novel findings regarding the

levels of client %CT that performed well in identifying subsequent client change. Client

%CT demonstrated optimal performance at rates of approximately 78%. This translates

to a ratio of nearly 4:1 between change talk and sustain talk, suggesting that the presence

of sustain talk is not in itself a sign of non-adherent MI, nor is the presence of change talk

a guarantee of treatment success, but the two variables should be viewed within the

context of all client speech to accurately indicated a client's likelihood of subsequent

change. Such a ratio could be useful for MI practitioners as an approximate indicator of

client motivation and could be easily monitored throughout a therapy session.

Clinical recommendations. There are specific treatment recommendations that may be useful when training and practicing motivational interviewing. Notably, a group of highly trained and skillful therapists showed average scores well below targeted integrity thresholds. Further, sub-threshold fidelity levels identified reliable and clinically significant changes in alcohol use. Without demonstrating expert recommended proficient or competent levels of treatment fidelity, therapists delivered effective MI. The sufficiency standards produced from the current research serve as a possible benchmark for measuring MI performance and signaling probable effectiveness, independent of expert recommendations. This may have an impact on how a trainer decides to structure an MI training and it may also have an impact on the amount of ongoing coaching necessary to bring a new MI practitioner up to a given performance criterion. For instance, a trainer or consultant may understand that higher %MIA and lower %OQ to be associated with effective MI and may spend more time focused on practicing MI Adherent statements and relatively less time practicing open questions. Feedback and coaching may center on encouraging MI practice that equally balances reflections and questions to reach the target R:Q ratio. A practice activity within MI training may acknowledge that effective MI can tolerate a small amount of MINA statements and focus practice on recognizing common situations when MINA statements occur or ways of providing frequent MIA statements.

The ROC curve analysis shows promising clinical applications for future research and practice. Given the straight-forward approach to creating and interpreting ROC curves, it may be possible within specific treatment settings to collect therapist MI fidelity measures along with client outcome variables and create unique ROC curves to

measure the level of treatment fidelity necessary to identify client change in a specific

setting, with a specific population, and for a specific targeted behavior change goal. This

would allow for an optimal tailoring of treatment to fit a unique treatment setting and

population. ROC curve analysis is specifically well suited for application within clinical

settings as it performs best when drawing a sample from a broad treatment population as

one would assume to encounter within a clinical setting made up of a heterogeneous

treatment seeking population.

It is important to note that the findings from this study came from a data set that

had been previously analyzed (Moyers et al., 2009). In that study, Moyers and colleagues

found that therapists who responded to client change talk with MI-consistent behaviors

were much more likely to have clients respond with *more* change talk and when

therapists responded to sustain talk with MI-consistent behaviors, clients were *less* likely

to respond with more sustain talk (Moyers et al., 2009). Moyers and colleagues also

found through multilevel analysis that higher levels of client change talk were associated

with fewer drinks per week at week-five of follow-up. Although this may not speak to

treatment adherence directly, it is consistent with the theory of effective MI-practice and

it speaks clearly to the use of specific therapeutic techniques in specific clinical instances.

It suggests an important finding from MI mechanisms research may be relevant when

discussing treatment adherence: effective MI may depend on more than just using the

correct behaviors, but using the correct behaviors in response to client speech. It may

matter more that the therapist responds with MIA behavior to a client's change talk than

it matters for the therapist to meet a higher threshold of %MIA. The importance of

moment-by-moment responding to client behavior may account for the findings from the

current study, which identified seemingly low fidelity levels as performing well in identifying meaningful client changes in alcohol use.

ROC curve analysis. Prior to this study, ROC curve analysis had been used widely as a method for decision making in medical diagnostics as well as psychological research, but it had seldom been applied to the setting of clinical psychology, specifically its utility as a decision making tool for identifying effective treatment adherence. To this end, the study findings suggest that ROC curve analysis may be used to identify optimal levels at which adherence variables most correctly identified subsequent client change. This provides a useful and promising methodology for establishing evidence-based fidelity thresholds for therapist treatment adherence.

Several practical considerations remain relevant to the future use of ROC curve analysis as a method of fidelity measurement. First, establishing an appropriate sample size for such analysis would provide answers to other clinicians and researchers who want to use this methodology in analyzing therapist treatment integrity. This study showed that the sample size of 118 was sufficient to detect an effect, but that the sample lacked power to compare the performance of one ROC curve with another. Future research in this area should use larger sample sizes that allow for the statistical comparison of different ROC curves. In addition, the availability of open source software capable of running ROC curve analysis would increase the likelihood that such analyses would be further explored in both clinical and research settings.

Conceptual issues remain unanswered regarding the use of ROC curve analysis in evaluating therapist treatment adherence. The ROC curve is an illustration of the relative performance of a diagnostic measure on identifying a given target behavior – in this case

treatment fidelity level identifying significant client changes in alcohol use and related

behaviors. Technically, an ideal result is to maximize true positive rate and minimize

false positive rate. Practically, it is important to decide what trade-off between sensitivity

and specificity is acceptable and useful within a given context. In terms of considering

therapist integrity level and client outcomes, it would be important to determine the

targeted true positive rate and an acceptable false positive rate. Given that therapy is an

endeavor with an acceptable level of false positivity – that is, the client may show little

change at times despite the therapist performing treatment with great fidelity –

identifying these levels of expected sensitivity and specificity would lend confidence to

the interpretation of ROC curve analysis results. A logical place to begin this process is

to establish base rates of client change within a treatment setting as foundational

performance thresholds. If ROC curve analysis is to be useful as a diagnostic

methodology in clinical settings in should – at minimum – show improvement upon the

existing base rate.

Another conceptual concern is the distance in terms of time and space that exists

between the indicator of therapist treatment integrity and the target of client change.

ROC curve analysis has been typically used in contexts where there is a shorter lapse

between the indicator and the target. For instance, in the context where ROC curve

analysis was developed, a radar blip indicated the target of an airplane. The indicator

was a sonar signal literally bounced off the target that it was detecting. In psychological

studies, ROC curve analysis has been used to identify perception thresholds, where the

indicator was awareness of perception and the target was the object perceived. In health

related fields, ROC curve analysis has been used in situations where a symptom-level of

a client functioned as an indicator of a target diagnosis. It is unclear if this represents a

confound in using ROC curve analysis for the purposes of this study or if it highlights the

robust nature of the methodology.

Relatedly, it is also important to note that ROC curve analysis is a methodology

for describing the performance of one variable on identifying the state of an outcome. It

is not a methodology for identifying causal relationships. The goal of the methodology is

to correctly identify the presence or absence of an outcome and to maximize the

performance of this identification rate. These results should be viewed as describing the

probability of a relationship rather than establishing a causal relationship.

Study limitations. One notable limitation of this study is the possibility that both

therapists and clients represented a skewed population and may not represent therapists

and clients in actual clinical settings. Therapists were highly trained and monitored

throughout their study involvement. It is possible that the MI that was observed in

Project MATCH was of higher quality than can be expected from uncontrolled

community settings. The client sample in this study was made up of individuals who

sought treatment for their alcohol use, freely consented to participate in a treatment study,

were randomly assigned to receive MET. This experience of treatment was different

from treatment as usual in that it was of shorter duration, no individuals were mandated

to treatment, and did not involve group therapy, which is a common component of

community-based treatment.

Motivational interviewing is understood as functioning from a combination of

relational and technical components. Fidelity thresholds in this study were primarily

focused on the technical aspects of MI: the questions that were asked, reflections that

were offered, specific   behaviors that were or we not in agreement with MI principles.

Therapist global behaviors were not available for this data sample and could not be

included when testing integrity thresholds. As a result, this study was unable to explore a

central aspect of MI, the way-of-being by which a therapist delivers specific therapeutic

language. It may be that a collection of therapeutic statements offered within a therapy

session function quite differently depending on whether they occur within a client-

centered, collaborative discussion focused towards a client-directed change or a context

in which a clinician explicitly and forcefully directs a client towards a change that the

therapist identifies as important. Such information about the therapeutic context might

have provided a useful session-wide characterization of the therapist's way-of-being.

This study was exploratory in nature and tested the utility of ROC analysis as a

method for measuring the performance of MI therapist fidelity levels on identifying client

change. As this was a novel methodology for evaluating the relationship between

treatment adherence and client outcomes, several variables were tested on the

identification of multiple client outcomes at various time points. Such combinations

resulted in multiple comparisons of fidelity measure, outcome, and time point. This

method of running multiple analyses was inconsistent with the body of ROC literature,

which most often tests the performance of one variable on identifying a dichotomous

outcome (typically disease status) or testing two ROC curves against one another to

determine the optimal performer. Given the large number of analyses, it the study results

were at risk for a type I error. The extent to which this issue impacted the observed

results was unclear as the statistic central to these analyses was not the p-value, but the

AUC value. Independent ROC analyses do not often report p-values and significance

testing with ROC analysis is used only when two ROC curves are compared against each

other. (Hanley & McNeal, 1982). Given the sample size for the current study, there was

not adequate power to compare paired ROC curves against one another (Bradley &

Longstaff, 2004).

A conservative methodology would have been to apply a Bonferroni correction to

the p-values of the analyses run. Such a correction would have rendered all findings

nonsignificant. This point bears further consideration. In this study, a p-value signifies

the probability that another sample of similar therapists and clients would result in equal

or more extreme results given the null hypothesis of no effect. This statistic, within the

context of ROC analysis, can be misleading as it speaks to the likelihood of similar

results with another sample, rather than the performance of the current sample. Further, a

Bonferroni corrected p-value would have been less than p-value = .001, which would

create an overly conservative p-value, protecting against a type I error and all but

ensuring a type II error. Additionally, given large enough sample size, an ROC analysis

with an AUC of .500 might result in a significant p-value, but a performance no different

from chance.

One possible solution was to use additional statistics, along with the p-value, to

evaluate the utility of the ROC analyses. Other statistics such as Youden's Index, the

AUC, and 95% confidence intervals provided meaningful measures of the utility of the

ROC analysis. Taken together, these measures help to demonstrate the extent to which

similar results could have been obtained by chance, the expected performance of the

observed data, and the point at which the selection variable exhibited optimal

performance. This information provides a comprehensive description of the data that cannot be drawn from *p*-values alone.

The limitations that arise from these repeated measures cannot be fully resolved within the current study. Future research can avoid such problems through a priori decisions about the client change threshold, outcome variable, and time point of interest. Such decisions may help to clarify the interpretation of statistical analysis, but beg other questions about how recovery is conceptualized. ROC analysis functions best when the performing variable is used to indicate a gold standard in outcome (such as clinical interview to diagnose mental health disorder or biopsy to diagnose cancer). Within the field of addictions, there still exists variability in how to define recovery, with most agreeing that recovery is a process with descriptive differences across time (White, 2007). Those interested in solving the above methodological problem will first have to address the conceptual issues of defining recovery at a particular point in the process.

Future directions. The use of ROC curve analysis has shown promise as a method for understanding the relationship between treatment fidelity and client outcomes. Given the broad application of MI, it may be important to establish context-specific fidelity standards that identify sufficient adherence to ensure effective treatment. Future research might explore the extending to which these findings generalize across different treatment settings, practitioners, and target behaviors.

Summary. This study, which considered both theoretical as well as empirical foundations of both MI treatment integrity and meaningful client change, and explored ROC curve analysis in a novel context, provided important suggestions towards the establishment of MI fidelity standards that are theoretically driven, empirically supported,

and practically useful in clinical contexts. The findings from this study suggest specific levels of therapist fidelity that can be used to identify significant client change with a level of accuracy that is significantly better than chance. These findings are of possible interest to those who would like to know the level of therapist treatment adherence necessary to provide the greatest likelihood of effective treatment and positive client change. Future research should further investigate such findings with larger sample populations and with different treatment seeking populations to explore the extent to which findings from this study can be generalized to other contexts.

References

Ang, D. C., Kaleth, A. S., Bigatti, S., Mazzuca, S., Saha, C., Hilligoss, J., Lengerich, M., & Bandy, R. (2011). Research to Encourage Exercise for Fibromyalgia (REEF): use of motivational interviewing design and method. *Contemporary Clinical Trials*, *32*(1), 59-68.

Baer, J. S., Carpenter, K. M., Beadnell, B., Stoner, S. A., Ingalsbe, M. H., Hartzler, B., & Drager, Z. (2012). Computer Assessment of Simulated Patient Interviews (CASPI): Psychometric Properties of a Web-Based System for the Assessment of Motivational Interviewing Skills. *Journal of Studies on Alcohol and Drugs, 73*(1), 154-164.

Bertholet, N., Faouzi, M., Gmel, G., Gaume, J., & Daeppen, J. B. (2010). Change Talk Sequence During Brief Motivational Intervention, Towards or Away From Drinking. *Addiction*, *105*(12), 2106-2112.

Bohman, B., Forsberg, L., Ghaderi, A., & Rasmussen, F. (2013). An Evaluation of Training in Motivational Interviewing for Nurses in Child Health Services. *Behavioural and Cognitive Psychotherapy, 41*(3), 329-343.

Burke, B. L., Arkowitz, H., & Menchola, M. (2003). The Efficacy of Motivational Interviewing: a Meta-analysis of Controlled Clinical Trials. *Journal of Consulting and Clinical Psychology*, *71*(5), 843.

Campbell, S. D., Adamson, S. J., & Carter, J. D. (2010). Client Language During Motivational Enhancement Therapy and Alcohol Use Outcome. *Behavioural and Cognitive Psychotherapy*, *38*(4), 399.

Carels, R. A., Darby, L., Cacciapaglia, H. M., Konrad, K., Coit, C., Harper, J., Versland,
     A. (2007). Using Motivational Interviewing As a Supplement to Obesity
     Treatment: A Stepped-Care Approach. *Health Psychology, 26*(3), 369-374.

Carpenter, K. M., Cheng, W. Y., Smith, J. L., Brooks, A. C., Amrhein, P. C., Wain, R.
     M., & Nunes, E. V. (2012). "Old Dogs" and New Skills: How Clinician
     Characteristics Relate to Motivational Interviewing Skills Before, During, and
     After Training. *Journal of Consulting and Clinical Psychology, 80*(4), 560-573.

Carroll, K. M., & Rounsaville, B. J. (2007). A Vision of the Next Generation of
     Behavioral Therapies Research in the Addictions. *Addiction*, *102*(6), 850-862.

Cimini, M. D., Martens, M. P., Larimer, M. E., Kilmer, J. R., Neighbors, C., &
     Monserrat, J. M. (2009). Assessing the Effectiveness of Peer-Facilitated
     Interventions Addressing High-Risk Drinking Among Judicially Mandated
     College Students. *Journal of Studies on Alcohol and Drugs*, 57-66.

Daeppen, J. B., Bertholet, N., Gmel, G., & Gaume, J. (2007). Communication During
     Brief Intervention, Intention to Change, and Outcome. *Substance Abuse*,*28*(3), 43-
     51.

D'Amico, E. J., Osilla, K. C., Miles, J. N. V., Ewing, B., Sullivan, K., Katz, K., &
     Hunter, S. B. (2012). Assessing Motivational Interviewing Integrity for Group
     Interventions with Adolescents. *Psychology of Addictive Behaviors, 26*(4), 994-
     1000.

DiClemente, C. C., & Velasquez, M. M. (2002). Motivational interviewing and the
     Stages of Change. *Motivational interviewing: Preparing people for change*, *2*,
     201-216.

Dunn, C., Deroo, L., & Rivara, F. P. (2001). The Use of Brief Interventions Adapted
    From Motivational Interviewing Across Behavioral Domains: A Systematic
    Review. *Addiction*, *96*(12), 1725-1742.

Evangeli, M., Longley, M., & Swartz, L. (2011). No Reductions and Some Improvements
    in South African Lay HIV/AIDS Counsellors' Motivational Interviewing
    Competence One Year After Brief Training. *Aids Care-Psychological and Socio-
    Medical Aspects of AIDS/HIV, 23*(3), 269-273.

Forsberg, L., Ernst, D., & Farbring, C. A. (2011). Learning Motivational Interviewing in
    a Real-life Setting: A Randomised Controlled Trial in the Swedish Prison Service.
    Criminal Behaviour and Mental Health, 21(3), 177-188.

Gallon, S. L., Ball, S. A., Ceperich, S., del Mar Garcia, M., Hamilton John, A., &
    Hausotter, W. (2006). Motivational Interviewing Assessment: Supervisory Tools
    for Enhancing Proficiency (MIA: STEP). *Kansas City, MO: Addiction
    Technology Transfer Center National Office*.

Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for
    Medical Diagnostic Test Evaluation. *Caspian Journal of Internal Medicine*,*4*(2),
    627.

Hanley, J. A., & McNeil, B. J. (1982). The Meaning and Use of the Area Under a
    Receiver Operating Characteristic (ROC) Curve. *Radiology*, *143*(1), 29-36.

Hannover, W., Thyrian, J. R., Roske, K., Grempler, J., Rumpf, H. J., John, U., & Hapke,
    U. (2009). Smoking Cessation and Relapse Prevention for Postpartum Women:
    Results From a Randomized Controlled Trial at 6, 12, 18 and 24 Months.
    Addictive Behaviors, 34(1), 1-8.

Hansen, A. B. G., Becker, U., Nielsen, A. S., Gronbk, M., & Tolstrup, J. S. (2012). Brief Alcohol Intervention by Newly Trained Workers Versus Leaflets: Comparison of Effect in Older Heavy Drinkers Identified in a Population Health Examination Survey: A Randomized Controlled Trial. *Alcohol and Alcoholism, 47*(1), 25-32.

Hettema, J., Steele, J., & Miller, W. R. (2005). Motivational interviewing. *Annual Review of Clinical Psychology*, *1*, 91-111.

Hogue, A., Dauber, S., Chinchilla, P., Fried, A., Henderson, C., Inclan, J., Reiner, R., & Liddle, H. A. (2008). Assessing Fidelity in Individual and Family Therapy for Adolescent Substance Abuse. *Journal of Substance Abuse Treatment, 35*(2), 137-147.

Houck, J. M., Moyers, T. B., Miller, W. R., Glynn, L. H., & Hallgren, K. A. (2010). Motivational Interviewing Skill Code (MISC) version2.5

Lane C, Huws-Thomas M, Hood K, Rollnick S, Edwards K, Robling M. (2005). Measuring Adaptations of Motivational Interviewing: The Development and Validation of the Behavior Change Counseling Index (BECCI). *Patient Education and Counseling* 2005;56:166-173.

Lindqvist, H., Forsberg, L. G., Forsberg, L., Rosendahl, I., Enebrink, P., & Helgason, A. R. (2013). Motivational Interviewing in an Ordinary Clinical Setting: A Controlled Clinical Trial at the Swedish National Tobacco Quitline. *Addictive Behaviors, 38*(7), 2321-2324.

Madson, M. B., Campbell, T. C., Barrett, D. E., Brondino, M. J., & Melchert, T. P. (2005). Development of the Motivational Interviewing Supervision and Training Scale. *Psychology of Addictive Behaviors*, *19*(3), 303.

Madson, M. B., Loignon, A. C., & Lane, C. (2009). Training in Motivational

      Interviewing: A Systematic Review. *Journal of Substance Abuse Treatment*,*36*(1),

      101-109.

Mastroleo, N. R., Mallett, K. A., Turrisi, R., & Ray, A. E. (2009). Psychometric

      Properties of the Peer Proficiency Assessment (PEPA): A Tool for Evaluation of

      Undergraduate Peer Counselors' Motivational Interviewing Fidelity. *Addictive*

      Behaviors, 34(9), 717-722.

McCambridge, J., Day, M., Thomas, B. A., & Strang, J. (2011). Fidelity to Motivational

      Interviewing and Subsequent Cannabis Cessation Among Adolescents. *Addictive*

      Behaviors, 36(7), 749-754.

Miller, W. R. (2000). Motivational Interviewing Skill Code (MISC): Coder's

      Manual. *University of New Mexico*.

Miller, W. R., & Johnson, J. L. (2001). When is it Motivational Interviewing?

      Addiction, 96(12), 1770-1771.

Miller, W. R., & Mount, K. A. (2001). A Small Study of Training in Motivational

      Interviewing: Does One Workshop Change Clinician and Client Behavior?

      Behavioural and Cognitive Psychotherapy, *29*(04), 457-471.

Miller, W. R., Moyers, T. B., Ernst, D., & Amrhein, P. (2009). *Motivational Interviewing*

      Skills Code (Version 2.1)

Miller, W. R., & Rollnick, S. (1991). *Motivational Interviewing: Preparing People to*

      Change Addictive Behavior. Guilford Press.

Miller, W. R., & Rollnick, S. (2009). Ten Things that Motivational Interviewing Is

      Not. *Behavioural and Cognitive Psychotherapy*, *37*(2), 129.

Miller, W. R., & Rollnick, S. (2012). *Motivational Interviewing: Preparing People for Change*. Guilford Press.

Miller, W. R., & Rollnick, S. (2014). The Effectiveness and Ineffectiveness of Complex Behavioral Interventions: Impact of Treatment Fidelity. *Contemporary Clinical Trials*, *37*(2), 234-241.

Miller, W. R., & Rose, G. S. (2009). Toward a Theory of Motivational Interviewing. *American Psychologist*, *64*(6), 527.

Miller, W. R., Yahne, C. E., Moyers, T. B., Martinez, J., & Pirritano, M. (2004). A Randomized Trial of Methods to Help Clinicians Learn Motivational Interviewing. *Journal of Consulting and Clinical Psychology*, *72*(6), 1050.

Moyers, T. B., Houck, J. M., Glynn, L. H., & Manuel, J. K. (2011). Can Specialized Training Teach Clinicians to Recognize, Reinforce, and Elicit Client Language in Motivational Interviewing? *Alcoholism: Clinical and Experimental Research, 35*(S1), 296.

Moyers, T.B., Manuel, J.K., & Ernst, D. (2014). *Motivational Interviewing Treatment Integrity Coding Manual 4.0*. Unpublished Manual.

Moyers, T. B., & Martin, T. (2006). Therapist Influence on Client Language During Motivational Interviewing Sessions. *Journal of Substance Abuse Treatment*,*30*(3), 245-251.

Moyers, T., Martin, T., Catley, D., Harris, K. J., & Ahluwalia, J. S. (2003). Assessing the Integrity of Motivational Interviewing Interventions: Reliability of the Motivational Interviewing Skills Code. *Behavioural and Cognitive Psychotherapy*, *31*(2), 177-184.

Moyers, T. B., Martin, T., Houck, J. M., Christopher, P. J., & Tonigan, J. S. (2009). From In-Session Behaviors to Drinking Outcomes: A Causal Chain for Motivational Interviewing. *Journal of Consulting and Clinical Psychology*, *77*(6), 1113.

Moyers, T. B., Martin, T., Manuel, J. K., Miller, W. R., & Ernst, D. (2010). Revised Global Scales: Motivational Interviewing Treatment Integrity 3.1. 1 (MITI 3.1. 1). *Unpublished Manuscript. Albuquerque: Center on Alcoholism, Substance Abuse and Addictions, University of New Mexico*.

Moyers, T. B., Martin, T., Manuel, J. K., Hendrickson, S. M., & Miller, W. R. (2005). Assessing Competence In The Use of Motivational Interviewing. *Journal of Substance Abuse Treatment*, *28*(1), 19-26.

Moyers, T. B., Miller, W. R., & Hendrickson, S. M. (2005). How Does Motivational Interviewing Work? Therapist Interpersonal Skill Predicts Client Involvement Within Motivational Interviewing Sessions. *Journal of Consulting and Clinical Psychology*, *73*(4), 590.

Noonan, W. C., & Moyers, T. B. (1997). Motivational Interviewing. *Journal of Substance Use*, *2*(1), 8-16.

Pollak, K. I., Alexander, S. C., Coffman, C. J., Tulsky, J. A., Lyna, P., Dolor, R. J., Ostbye, T. (2010). Physician Communication Techniques and Weight Loss in Adults Project CHAT. *American Journal of Preventive Medicine, 39*(4), 321-328. doi: 10.1016/j.amepre.2010.06.005

Project MATCH Research Group. (1998). Matching Alcoholism Treatments to Client Heterogeneity: Treatment Main Effects and Matching Effects on Drinking During Treatment. *Journal of Studies in Alcohol*, *59*(6), 631-639.

Rollnick, S. (2001). Enthusiasm, Quick Fixes, and Premature Controlled Trials.

     Addiction, 96, 1769–1770.

Schmiege, S. J., Broaddus, M. R., Levin, M., & Bryan, A. D. (2009). Randomized Trial

     of Group Interventions to Reduce HIV/STD Risk and Change Theoretical

     Mediators Among Detained Adolescents. *Journal of Consulting and Clinical*

     Psychology, 77(1), 38-50.

Smith, J. L., Carpenter, K. M., Amrhein, P. C., Brooks, A. C., Levin, D., Schreiber, E. A.,

     Travaglini, L., Hu, M., & Nunes, E. V. (2012). Training Substance Abuse

     Clinicians in Motivational Interviewing Using Live Supervision Via

     Teleconferencing. *Journal of Consulting and Clinical Psychology, 80*(3), 450-

     464.

Sobell, M. B., & Sobell, L. C. (1996). *Problem drinkers: Guided Self-Change Treatment*.

     Guilford Press.

Thush, C., Wiers, R. W., Moerbeek, M., Ames, S. L., Grenard, J. L., Sussman, S., &

     Stacy, A. W. (2009). Influence of Motivational Interviewing on Explicit and

     Implicit Alcohol-Related Cognition and Alcohol Use in At-Risk Adolescents.

     Psychology of Addictive Behaviors, 23(1), 146-151.

Tollison, S. J., Lee, C. M., Neighbors, C., Neil, T. A., Olson, N. D., & Larimer, M. E.

     (2008). Questions and Reflections: The Use of Motivational Interviewing

     Microskills in a Peer-Led Brief Alcohol Intervention for College Students.

     Behavior Therapy, 39(2), 183-194.

Turrisi, R., Larimer, M. E., Mallett, K. A., Kilmer, J. R., Ray, A. E., Mastroleo, N. R.,

     Geisner, I., & Montoya, H. (2009). A Randomized Clinical Trial Evaluating a

Combined Alcohol Intervention for High-Risk College Students. *Journal of Studies on Alcohol and Drugs, 70*(4), 555-567.

Vader, A. M., Walters, S. T., Prabhu, G. C., Houck, J. M., & Field, C. A. (2010). The Language of Motivational Interviewing and Feedback: Counselor Language, Client Language, and Client Drinking Outcomes. *Psychology of Addictive Behaviors*, *24*(2), 190.

White, W. L. (2007). Addiction Recovery: Its Definition and Conceptual Boundaries. *Journal of Substance Abuse Treatment*, *33*(3), 229-241.

Tables and figures

Table 1
MITI 3.0 Adherence thresholds for therapist MI performance.

Category	"Competency"	"Proficiency"
Percent Complex Reflections	50%	40%
Percent Open Questions	70%	50%
Reflection to Question Ratio	2.0	1.0
Percent MI-Adherence	100%	90%

Table 2
Descriptive statistics of client drinking outcomes.

Time point	PDA*	DDD*	DrInC
Baseline	.41 (.31)	17.47 (11.90)	54.27 (23.97)
1 week	.86 (.30)	2.79 (7.36)	---
2 week	.89 (.27)	3.37 (8.41)	---
3 month	.89 (.26)	2.72 (5.33)	25.72 (33.43)
6 month	.86 (.25)	5.74 (9.32)	24.58 (30.57)

--- = Data unavailable.
* Mean scores for PDA and DDD include cases where no drinking was reported.

Table 3
Percentage of reliable change, functional change, and clinically significant change for client outcome variables.

Time point	PDA			DDD			DrInC		
	%RC	%FC	%CSC	%RC	%FC	%CSC	%RC	%FC	%CSC
2 wk.	76.07	38.46	38.46	48.72	15.38	15.38	--	--	--
3 mo.	75.42	43.22	43.22	50.85	16.95	16.95	45.33	34.67	35.14
6 mo.	73.73	34.75	34.75	42.37	13.56	13.56	46.67	32.00	33.33

-- = Data not available.

Table 4

Descriptive statistics for therapist and client in-session behaviors from session one

Therapist variable	Mean (s.d.)
%CR	.263 (.136)
%OQ	.241 (.120)
%MIA	.734 (.181)
R:Q	1.01 (.479)
Complex reflections	16.228 (11.474)
Simple reflections	44.847 (20.372)
Open questions	15.033 (8.528)
Closed questions	51.949 (27.090)
MIA	18.372 (10.352)
MINA	7.550 (10.534)

Client variable	Mean (s.d.)
Client %CT	.755 (.128)
Client change talk	54.872 (24.340)
Client sustain talk	18.661 (14.075)

* = Independent samples T-test of differences between sample means, p value < .05
%CR = % complex reflections, %OQ = % open questions, %MIA = % MI adherent behavior, R:Q = Ratio of reflections to questions, MIA = MI adherent behaviors, MINA = MI nonadherent behaviors, %CT = % change talk

Table 5

Results from ROC curve analysis of therapist fidelity level on clinically significant change in drinks per drinking day (DDD) at various time points.

Change Threshold	Fidelity Variable	Outcome Variable	AUC (95% C.I.)
CSC	%OQ	2 wk. DDD	.701 (.573 – .828)*
	%CR	2 wk. DDD	.435 (.298 – .573)
	%MIA	2 wk. DDD	.669 (.542 – .796)**
	R:Q	2 wk. DDD	.663 (.532 – .794)**
	%OQ	3 mo. DDD	.627 (.496 – .758)
	%CR	3 mo. DDD	.506 (.364 – .647)
	%MIA	3 mo. DDD	.626 (.505 – .747)
	R:Q	3 mo. DDD	.607 (.496 – .758)
	%OQ	6 mo. DDD	.654 (.513 – .795)**
	%CR	6 mo. DDD	.527 (.374 – .680)
	%MIA	6 mo. DDD	.639 (.512 – .765)
	R:Q	6 mo. DDD	.613 (.469 – .757)

*p-value < .01, **p-value < .05

Table 6

Results from ROC curve analysis of therapist fidelity level on clinically significant change in percent days abstinent (PDA) at various time points

Change Threshold	Fidelity Variable	Outcome Variable	AUC (95% C.I.)
CSC	%OQ	2 wk. PDA	.546 (.439 – .653)
	%CR	2 wk. PDA	.433 (.322 – .543)
	%MIA	2 wk. PDA	.588 (.483 – .693)
	R:Q	2 wk. PDA	.507 (.398 – .616)
	%OQ	3 mo. PDA	.512 (.406 – .618)
	%CR	3 mo. PDA	.424 (.318 – .529)
	%MIA	3 mo. PDA	.558 (.453 – .662)
	R:Q	3 mo. PDA	.523 (.417 – .628)
	%OQ	6 mo. PDA	.536 (.427 – .645)
	%CR	6 mo. PDA	.483 (.370 – .597)
	%MIA	6 mo. PDA	.589 (.484 – .694)
	R:Q	6 mo. PDA	.525 (.427 – .645)

Table 7
Results from ROC curve analysis of therapist fidelity level on clinically significant change in alcohol related consequences (DrInC scores) at various time points

Change Threshold	Fidelity Variable	Outcome Variable	AUC (95% C.I.)
CSC	%OQ	3 mo. DrInC	.544 (.410 – .679)
	%CR	3 mo. DrInC	.543 (.399 – .686)
	%MIA	3 mo. DrInC	.565 (.426 – .704)
	R:Q	3 mo. DrInC	.440 (.300 – .581)
	%OQ	6 mo. DrInC	.604 (.472 – .737)
	%CR	6 mo. DrInC	.534 (.394 – .674)
	%MIA	6 mo. DrInC	.588 (.447 – .729)
	R:Q	6 mo. DrInC	.456 (.312 – .599)

Table 8
Results from ROC curve analysis of therapist fidelity level on reliable change in drinks per drinking day (DDD) at various time points

Change Threshold	Fidelity Variable	Outcome Variable	AUC (95% C.I.)
RC	%OQ	2 wk. DDD	.605 (.503 – .708)
	%CR	2 wk. DDD	.455 (.348 – .561)
	%MIA	2 wk. DDD	.584 (.480 – .687)
	R:Q	2 wk. DDD	.579 (.503 – .708)
	%OQ	3 mo. DDD	.597 (.494 – .700)
	%CR	3 mo. DDD	.485 (.380 – .590)
	%MIA	3 mo. DDD	.561 (.457 – .665)
	R:Q	3 mo. DDD	.554 (.449 – .658)
	%OQ	6 mo. DDD	.626 (.524 – .727)**
	%CR	6 mo. DDD	.500 (.394 – .606)
	%MIA	6 mo. DDD	.604 (.501 – .707)
	R:Q	6 mo. DDD	.526 (.421 – .631)

*p-value < .01, **p-value < .05

Table 9

Results from ROC curve analysis of therapist fidelity level on reliable change in percent days abstinent (PDA) at various time points

Change Threshold	Fidelity Variable	Outcome Variable	AUC (95% C.I.)
RC	%OQ	2 wk. PDA	.570 (.441 – .700)
	%CR	2 wk. PDA	.447 (.324 – .571)
	%MIA	2 wk. PDA	.645 (.525 – .765)**
	R:Q	2 wk. PDA	.517 (.396 – .638)
	%OQ	3 mo. PDA	.523 (.395 – .651)
	%CR	3 mo. PDA	.479 (.358 – .600)
	%MIA	3 mo. PDA	.627 (.495 – .758)**
	R:Q	3 mo. PDA	.546 (.430 – .661)
	%OQ	6 mo. PDA	.600 (.485 – .716)
	%CR	6 mo. PDA	.497 (.375 – .619)
	%MIA	6 mo. PDA	.639 (.523 – .755)**
	R:Q	6 mo. PDA	.569 (.452 – .686)

*p-value < .01, **p-value < .05

Table 10

Results from ROC curve analysis of therapist fidelity level on reliable change in alcohol related consequences (DrInC scores) at various time points

Change Threshold	Fidelity Variable	Outcome Variable	AUC (95% C.I.)
RC	%OQ	3 mo. DrInC	.530 (.397 – .664)
	%CR	3 mo. DrInC	.550 (.417 – .683)
	%MIA	3 mo. DrInC	.582 (.452 – .712)
	R:Q	3 mo. DrInC	.515 (.383 – .647)
	%OQ	6 mo. DrInC	.631 (.505 – .757)
	%CR	6 mo. DrInC	.500 (.366 – .634)
	%MIA	6 mo. DrInC	.610 (.482 – .738)
	R:Q	6 mo. DrInC	. 528 (.395 – .661)

Table 11
Characteristics of optimal treatment integrity levels for statistically significant ROC
curve analyses of fidelity level identifying client change

Change Index	Fidelity Variable	Outcome Variable	Optimal Cutoff	TPR	FPR	Youden's Index Score
CSC	%OQ	2 wk. DDD	30.82 %	.611	.182	.429
	%MIA	2 wk. DDD	77.04 %	.722	.424	.298
	R:Q	2 wk. DDD	0.927	.778	.404	.374
	%OQ	6 mo. DDD	30.82 %	.563	.196	.367
RC	%OQ	6 mo. DDD	30.62%	.380	.162	.218
	%MIA	2 wk. PDA	71.01%	.652	.393	.259
	%MIA	3 mo. PDA	70.29%	.697	.379	.318

Table 12
Results from ROC curve analysis of percent client change talk on reliable change in
drinks per drinking day (DDD), percent days abstinent (PDA) and alcohol related
consequences (DrInC scores) at various time points

Change Threshold	Outcome Variable	AUC (95% C.I.)
CSC	2 wk. DDD	.636 (.523 – .749)
	3 mo. DDD	.589 (.475 – .703)
	6 mo. DDD	.621 (.496 – .747)
	2 wk. PDA	.529 (.421 – .637)
	3 mo. PDA	.486 (.379 – .593)
	6 mo. PDA	.564 (.455 – .673)
	3 mo. DrInC	.669 (.533 – .805)**
	6 mo. DrInC	.663 (.524 – .802)**
RC	2 wk. DDD	.525 (.419 – .630)
	3 mo. DDD	.505 (.399 – .610)
	6 mo. DDD	.500 (.395 – .606)
	2 wk. PDA	.592 (.470 – .713)
	3 mo. PDA	.541 (.421 – .662)
	6 mo. PDA	.548 (.427 – .669)
	3 mo. DrInC	.694 (.572 – .816)**
	6 mo. DrInC	.666 (.543 – .790)**

*p-value < .01, **p-value < .05

Table 13

Characteristics of optimal percent change talk levels for statistically significant ROC curve analyses of client percent change talk identifying client change.

Change Index	Outcome Variable	Optimal Cutoff	TPR	FPR	Youden's Index Score
CSC	3 mo. DrInC	78.69%	.654	.313	.341
	6 mo. DrInC	78.69%	.680	.340	.340
RC	3 mo. DrInC	78.69%	.647	.268	.379
	6 mo. DrInC	72.00%	.800	.475	.325
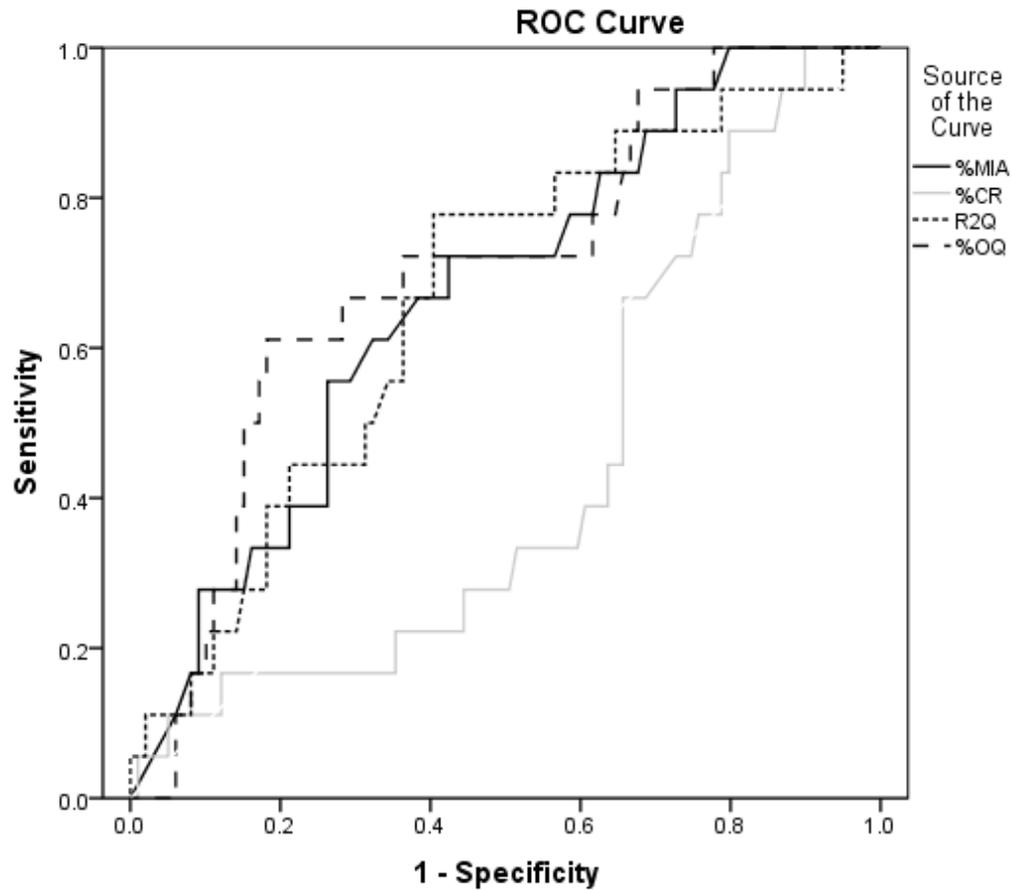
Figure 1
ROC curve of multiple therapist session-one adherence variables on identifying clinically significant change in drinks per drinking day at two-weeks of treatment.
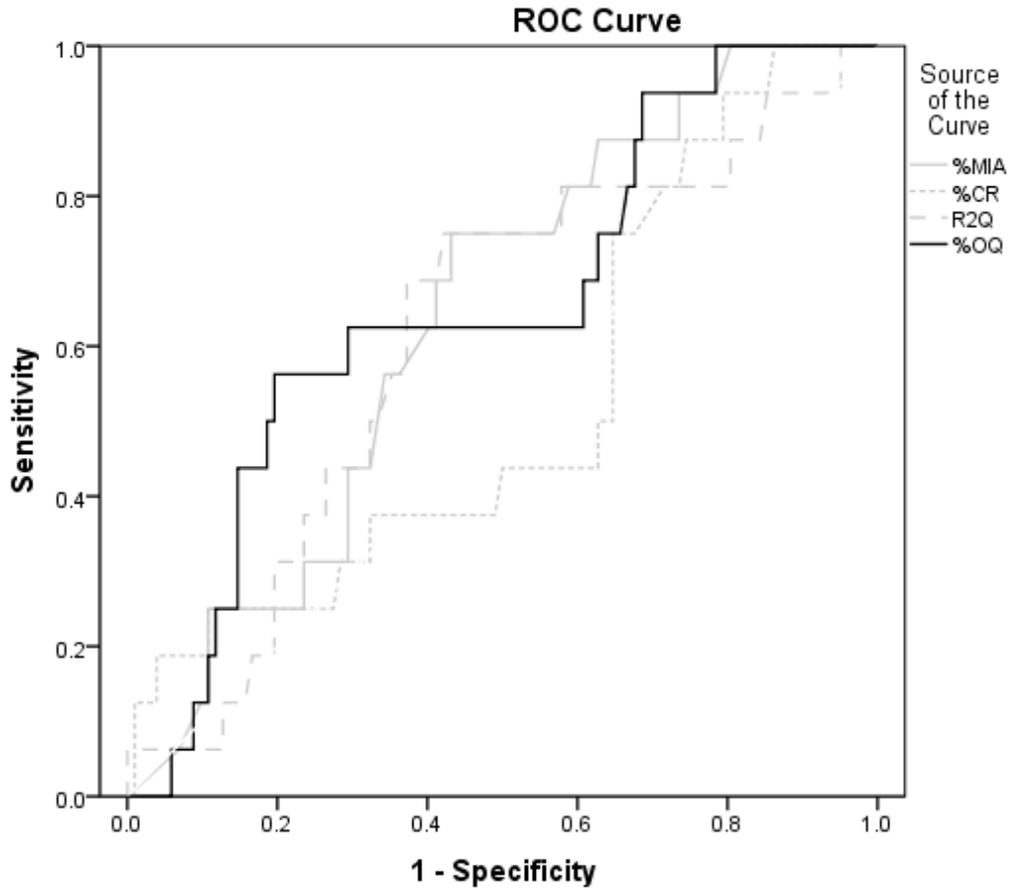
Figure 2
ROC curve of multiple therapist session-one adherence variables on identifying clinically significant change in drinks per drinking day at six-months of treatment.
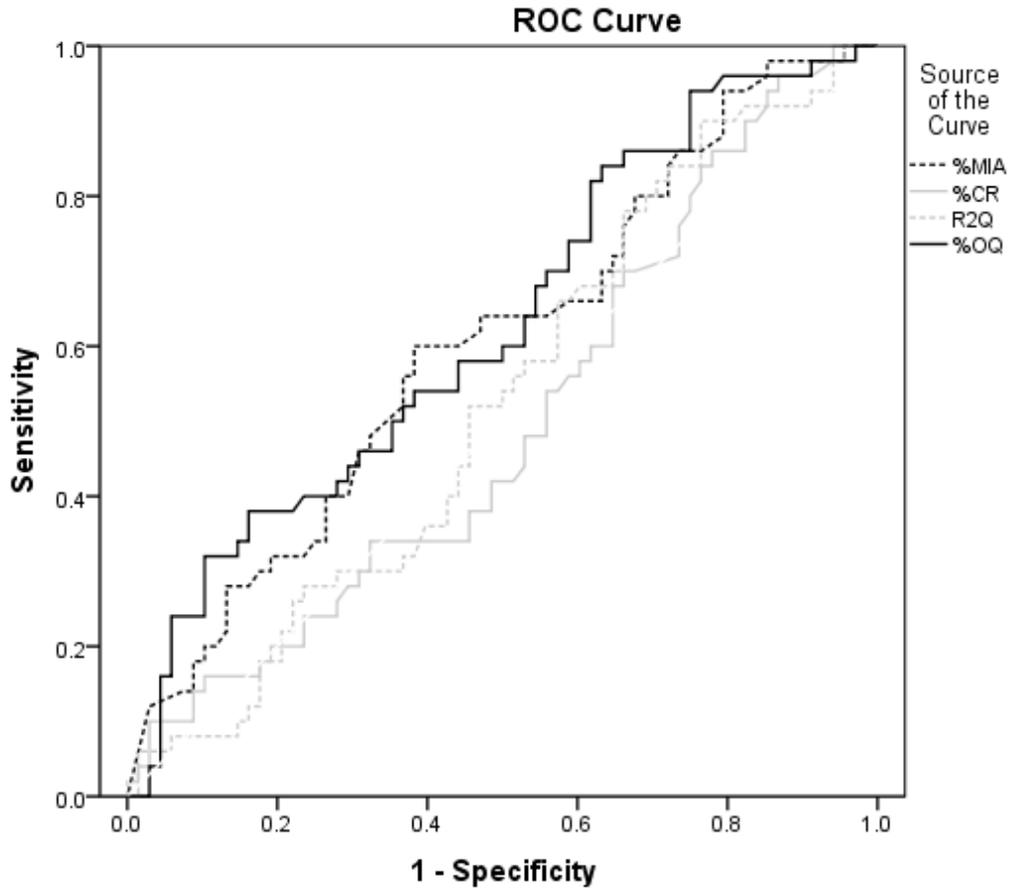
Diagonal segments are produced by ties.

Figure 3
ROC curve of multiple therapist session-one adherence variables on identifying reliable change in drinks per drinking day at six-months of treatment.
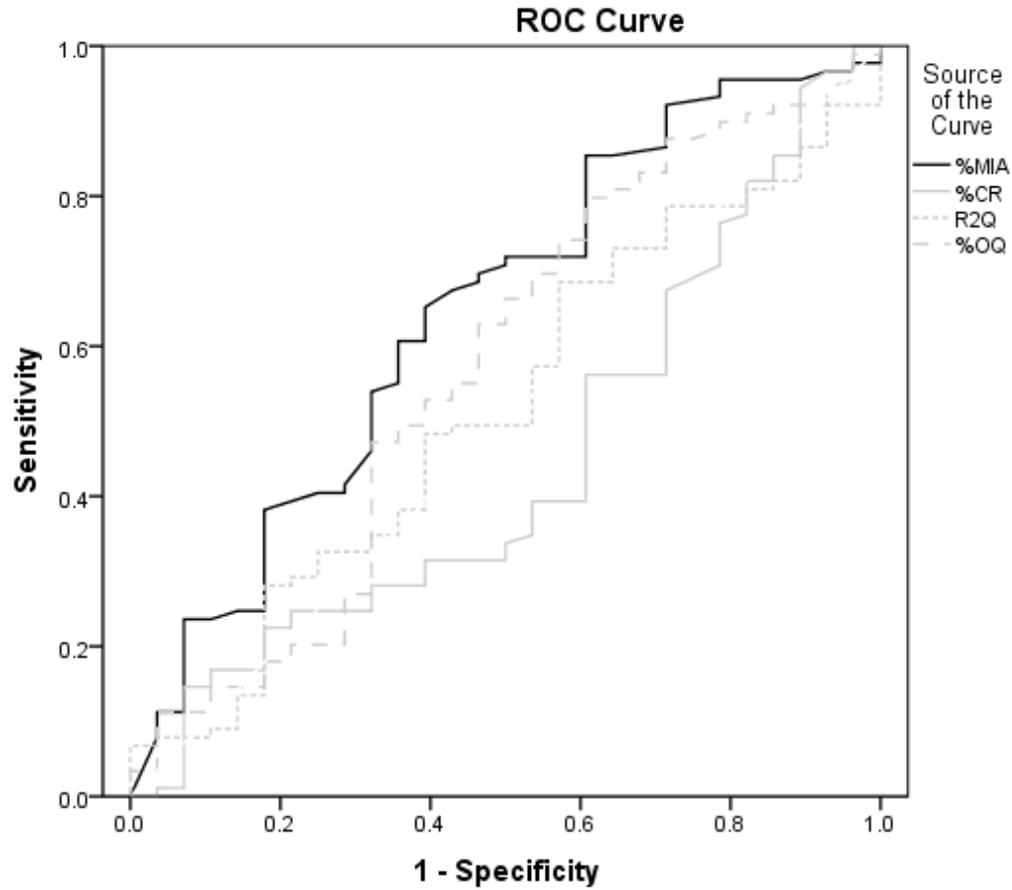
Diagonal segments are produced by ties.

Figure 4
ROC curve of multiple therapist session-one adherence variables on identifying reliable change in percent days abstinent at two-weeks of treatment.
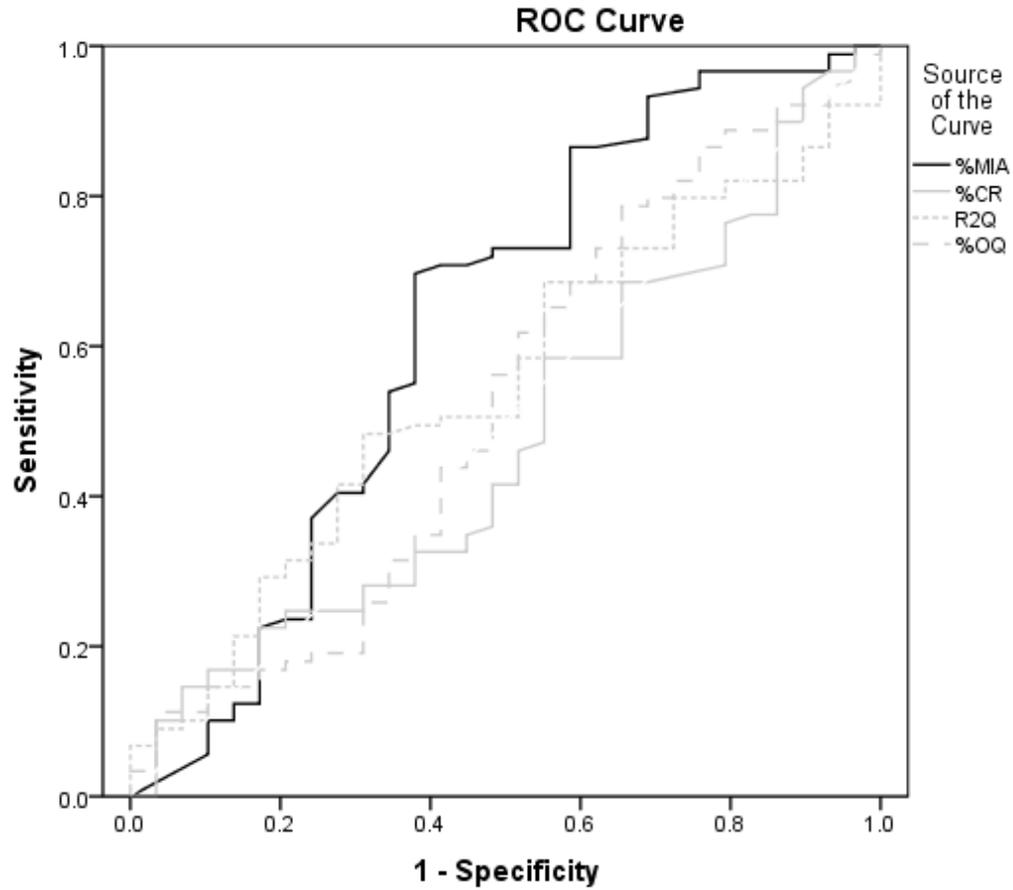
Diagonal segments are produced by ties.

Figure 5
ROC curve of multiple therapist session-one adherence variables on identifying reliable change in percent days abstinent at three-months of treatment.
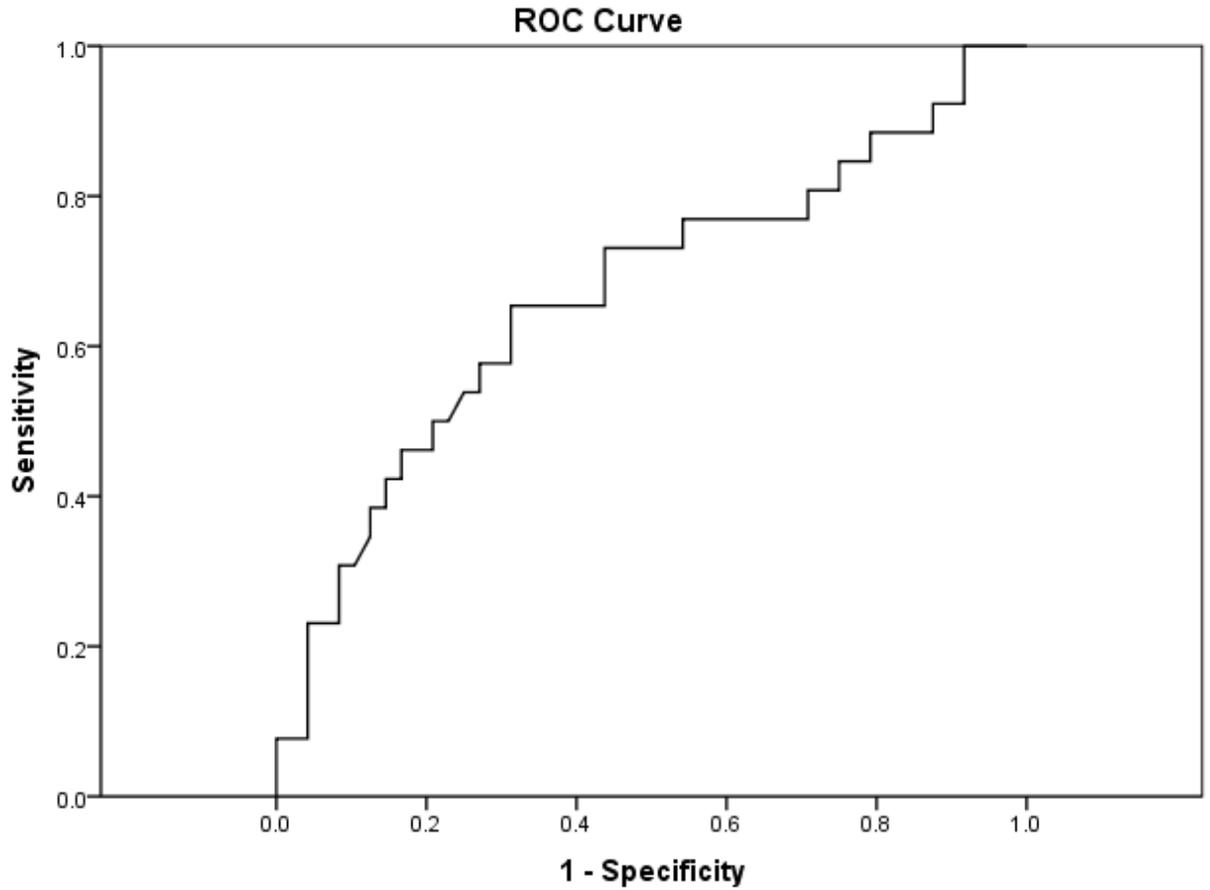
Figure 6
ROC curve of client session-one percent change talk on identifying clinically significant change in alcohol-related consequences (DrInC scores) at three-months of treatment.
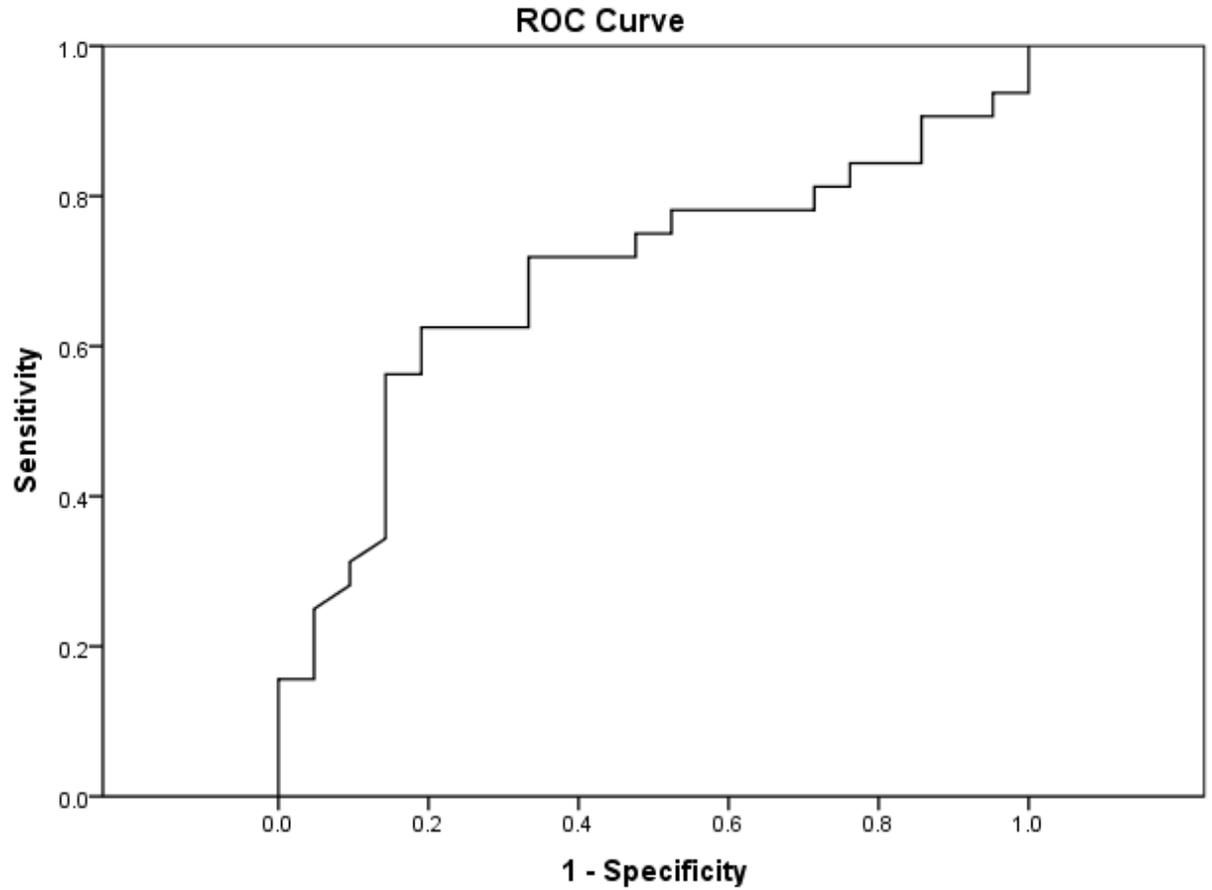
Diagonal segments are produced by ties.

Figure 7
ROC curve of client session-one percent change talk on identifying reliable change in alcohol-related consequences (DrInC scores) at three-months of treatment.