Contemporary Clinical Trials 37 (2014) 234–241

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

The effectiveness and ineffectiveness of complex behavioral interventions: Impact of treatment fidelity

William R. Miller [a,*], Stephen Rollnick [b]

[a] The University of New Mexico, United States
[b] Cardiff University, Wales, United Kingdom

ARTICLE INFO

Article history:
Received 1 November 2013
Revised 13 January 2014
Accepted 18 January 2014
Available online 24 January 2014

Keywords:
Clinical trials
Quality assurance
Fidelity
Training
Clinical trials
Motivational interviewing

ABSTRACT

There is often wide variability in the reported effects of complex behavioral interventions. Effectiveness can vary across studies, sites, and providers. A factor that has been insufficiently considered is the fidelity of the behavioral treatment that was provided. Low quality practice could be likened to partial doses of a vaccine or antibiotic: the right idea but insufficient strength. Using motivational interviewing (MI) as an example, the authors describe three quality conditions that should be present for a study to be regarded as a trial of a complex behavioral intervention: (1) The treatment should clearly contain the components that are theoretically or empirically related to its efficacy; (2) providers should be trained to an adequate and specified criterion of proficiency before treating trial patients; and (3) the fidelity of treatment should be documented by reliable coding of practice throughout the study and reported in a manner that permits comparison with skill levels in other trials. The authors also discuss bona fide intervention failures despite strong clinical trial methodology, offering recommendations for future outcome research.

1. Introduction

Mixed findings are common in clinical trials of behavioral, pharmacological, and medical interventions. Whereas pharmaceutical trials can maintain tight control over the content and dosage of medications delivered, complex behavioral interventions pose special challenges. It is impossible to include a true double-blind placebo condition in which both patient and provider are unaware of the specific treatment being delivered. For many years behavioral trials simply named and briefly described the treatment that was allegedly provided. Detailed therapist manuals were introduced to standardize and better specify treatments, but the presence of a manual does not indicate how well it was followed. The actual implementation of evidence-based treatments is a complex research topic in itself [1].

An example of the complexities of treatment evaluation is provided by research on motivational interviewing (MI), a collaborative counseling style for strengthening a person's own motivation and commitment to change [2]. Interventions identified as MI have been evaluated in hundreds of outcome studies including over 200 randomized clinical trials across a broad range of problem areas [3]. Meta-analyses have reported small to medium effects on average with highly variable effect sizes [4–6], and have concluded significant efficacy for substance use [7–9], smoking [10], weight loss [11], gambling [12], and medical outcomes such as blood pressure, cholesterol, dental caries, HIV viral load, and mortality [13].

Embedded within these averages are numerous trials finding no significant effect of MI-based training and interventions, including some in which we, as the original developers of MI, participated [14,15]. In one meta-analysis [6], the mean effect size across a wide range of health problems was $g = .22$ ($p < .001$), and only 77 of 132 studies (58%) yielded effect sizes

* Corresponding author.
E-mail address: WRMILLER@UNM.EDU (W.R. Miller).

1551-7144/$ – see front matter
http://dx.doi.org/10.1016/j.cct.2014.01.005

of .20 or greater, with wide variability ranging up to $g = 2.06$, which was not a statistical outlier.

In this article we consider reasons why behavioral interventions such as MI may be found to be ineffective, focusing in particular on treatment implementation and fidelity issues [16]. We conclude our discussion with recommendations for future clinical research and training.

2. Intervention content

In contrast to pharmacotherapies, the title given to a complex behavioral intervention conveys very little about what was actually delivered. Even the intended content of treatment can vary widely within a domain such as "cognitive–behavior therapy." Only a few behavioral interventions have been trademarked or patented in an attempt to control content, and even for these the nature of the intended treatment in a particular clinical trial needs to be specified.

One factor that can influence outcomes, then, is the specific content of the treatment that was meant to be delivered in a clinical trial. Complex behavioral interventions may contain components with specific efficacy, but also superstitious elements that have no effect or even detrimental impact on treatment outcome. A starting point is to ensure that the intervention contains what are theoretically expected to be active ingredients [16]. Research on mechanisms of action may provide evidence for empirically-supported components that should be present, yielding greater clarity about what aspects can be adapted (e.g., across cultures) without compromising efficacy.

The content delivered has varied widely in trials of "MI," which has been confused with other types of interventions including cognitive–behavior therapy, client-centered counseling, and decisional balance [12,17,18]. Some interventions that have been labeled as MI seem to bear little resemblance to the method that we developed. For example, Kuchipudi and colleagues [19] found no effect of a 3-session "MI" intervention (relative to treatment as usual) for patients with pancreatitis, ulcer or cirrhosis who had not responded to prior advice to decrease their drinking. The intervention consisted of "interviews with three different persons emphasizing the need for and benefits of alcoholism therapy. The person's health and drinking were reviewed from the viewpoint and with the authority of the director of the unit" (p. 357). Such a didactic, persuasive, and authoritarian approach sounds entirely contrary to the style of MI. Because we decided not to trademark or copyright the name of MI, it can be attached to methods unrelated or even antithetical to our original approach. This has contributed to confusion regarding the basic nature of MI [20]. From our perspective [2] the fundamental characteristics of MI are:

1. A person-centered non-authoritarian counseling style as originally described by Carl Rogers and his students [21,22], and
2. A clearly identified change goal toward which the conversation is directed, and
3. Differential evoking and strengthening of the person's own motivations for change.

One possible reason for variability of effects in clinical trials, then, is diversity in what was meant to be delivered.

3. Quality assurance

The behavioral intervention that was intended to be tested, even if well specified, may bear little resemblance to what was actually provided if there is inadequate quality assurance. In an early trial of the community reinforcement approach for alcohol dependence [23], the investigators (including the first author — WM) prepared a manual, trained therapists in the intended treatment, and met with them weekly to discuss cases. All sessions were recorded, but the audiotapes were not coded for fidelity until the treatment phase had ended. On examination the sessions contained shockingly little of the intended procedures. "We know now," they reported, "that what therapists say they do may bear little resemblance to what is actually done, even when sessions are recorded" (p. 103).

A common limitation, particularly in earlier trials of MI, was a lack of documentation as to what intervention was actually delivered. Rollnick [24] commented that, "Put bluntly, we do not know what went on inside the consultation in most of the studies. Why researchers and funding bodies overlooked this requirement in many studies is puzzling" (pp. 1769–1770). Sometimes there has been little more than a statement that "MI techniques were used."

When the intervention was well characterized and consistent with a MI approach, this is still no guarantee that the treatment was delivered as described. Even with a detailed therapist manual, the delivery and outcomes of MI often vary widely across providers [25,26]. In one meta-analysis [4] the effect size of MI was twice as large in studies without a therapist manual, as compared with those using a manual to standardize the intervention.

In the absence of quality assurance (QA) that is based on recordings of actual practice, it is impossible to determine what behavioral intervention was provided. Many negative trials of "MI" included no QA measure at all. In others the QA procedure was inadequate to determine whether MI had been delivered. A large trial with children with type 1 diabetes and their caregivers [15] evaluated an intervention that "was drawn from, but did not fully equate to, motivational interviewing" and found no beneficial effect. The trial carefully documented outcomes but unfortunately relied on novel untested global rating scales for QA of the intervention. Consequently it was difficult to determine what MI skill level was attained and practiced by providers in the trial.

Well-validated QA tools are available to provide measures of MI fidelity with inter-rater reliability [27] that do change with training and predict treatment outcome [28–35]. One widely used system is the Motivational Interviewing Treatment Integrity (MITI [36,37]) code containing both global rating scales and provider behavior counts for which normative values and recommended competence thresholds are available. The MI Skills Code [38,39], from which the MITI was derived, also quantifies client responses during MI sessions, allowing the linkage of therapist and client process measures with clinical outcomes [40–42].

4. Intervention fidelity

People cannot benefit from a treatment to which they have not been exposed [1]. No fixed dose of initial training is likely to

produce consistent practitioner fidelity. Ongoing feedback and coaching generally improve the quality of behavioral interventions [43] and of MI in particular [44], and the amount of such training that is necessary varies across providers. Another potential source of ineffectiveness, then, is insufficient initial training to establish and maintain intervention fidelity. Broekhuizen and colleagues [45] relied on a 3-day training in preparing lifestyle coaches to deliver MI in a large trial to reduce LDL cholesterol, using a well-validated instrument to assess intervention fidelity. The intervention yielded no significant effect on outcomes, but none of analyzed counseling sessions met thresholds for MI fidelity, and fewer than half of patients had received the intended intervention. In a multisite trial finding no effect of MI [46], "fidelity was monitored and maintained through direct observation, review of a random sample of audiotapes of counsellor-participant sessions, and review of case notes," (p. 175) but the QA review procedure was unspecified and no data were provided to document fidelity.

An ambiguity in QA monitoring is that it is unclear what level of MI fidelity is "good enough" to facilitate change within particular contexts or sufficient to conclude that MI was actually delivered and tested. Published provisional competence thresholds (such as an average of 3.5 on 5-point global rating scales [2]) have not been empirically derived. The unanswered issue here is analogous to dose–response analyses in pharmacological research.

Clinical trials should specify the expected and actual attained levels of therapist proficiency in delivering a behavioral treatment. What proficiency threshold was required to certify therapists prior to treating trial patients, and what fidelity level obtained in treatment delivery?

5. Adequacy of treatment

An issue often insufficiently considered or reported in behavioral trials is how to handle sub-threshold therapist performance. When guidelines are specified, they often involve procedures for de-certifying (red-lining) providers to not treat any further trial cases until remedial steps have been completed and performance meets or exceeds the trial's standards. This is a sensible procedure, but depending on the rapidity and reliability of tape review, decertification can lag far behind drift in fidelity.

And how should cases be handled when the delivered treatment quality fell below trial standards? One option is to exclude such cases from analysis to ensure that what is being evaluated is good-quality treatment. This introduces worrisome bias into a trial, however, because treatment fidelity varies within as well as between therapists, and may be related to factors such as patient characteristics. Thus selectively excluding low-fidelity cases changes the sample in unpredictable ways. We recommend retaining such cases and analyzing outcome data both ways: for all treated cases, and for high-fidelity cases.

Another issue of treatment adequacy is dosage. With an active treatment, dose is presumably a relevant variable. Is there a minimum length of time, number of sessions, or set of procedures to which participants should be exposed in order to be regarded as treated? Here there is a clear research precedent: to report outcomes for both the full intent-to-treat sample and for patients regarded as treated. In addition to the absolute amount of treatment delivered, the spacing or scheduling of "dosage" may be an important consideration.

6. Bona fide intervention failures

Most perplexing are studies where all of the previously described methodological issues were adequately addressed, and nevertheless the intervention failed to produce benefit. For example, the clinical method of MI was apparently well understood, quality assurance procedures were in place, and fidelity of treatment delivery was monitored with reliable methods and found to be good. Short of other compromising factors such as insufficient statistical power to detect an effect, these can reasonably be interpreted as failures of the intervention itself. Such failures, in combination with other positive trials of the same intervention with similar populations and target problems, point to insufficiently understood (or controlled) determinants of efficacy.

6.1. Single-site efficacy studies

An example of such a negative trial was conducted in our clinic in New Mexico [14] treating people with drug use disorders (primarily stimulants and opiates). The therapists were trained and supervised by the first author (WM) in manual-guided delivery of one-session motivational enhancement therapy (MET — a combination of MI with assessment feedback). All sessions were recorded and high therapist fidelity was documented using the MI Skills Code. Outcomes were quantified through timeline follow-back interviews [47,48] and at no point during 12 months of follow-up did we observe any intervention effect.

Psycholinguistic analyses in this study compared in-session speech patterns of clients who had more favorable (abstinence) versus less favorable outcomes (continued drug use) [49]. Those with better outcomes had shown a steady increase in change talk (strength of commitment language for drug abstinence) over the course of the MET session. A very different pattern of rising and falling commitment was found for clients with less favorable outcomes. Because the intervention had been standardized by a therapist manual, the drops in client commitment could be identified as occurring at two particular points in the sequence of the session: during assessment feedback, and when a change plan was initiated. In retrospect, the manual inadvertently violated the clinical style of MI by requiring therapists to continue assessment feedback and to complete a change plan whether or not the client was responding to it well. One cannot know whether this inflexibility accounted for failure of the intervention, but process analyses linking in-session speech to adverse outcomes are consistent with this flaw. The fact that the manual was written by the first author (WM) illustrates how readily seemingly small variations in delivery may have unintended consequences.

6.2. Multisite efficacy trials

The design logic of multisite trials is to amass a large sample across multiple recruitment sites and thereby obtain an estimate of effect that is not rooted in the vicissitudes of a particular local context. If the study is conducted in ongoing community clinics and provided by regular agency staff, a

multisite trial begins to bridge the gap from efficacy to effectiveness, and that has been the normative design for studies conducted within the National Institute on Drug Abuse Clinical Trials Network (CTN), a collaboration involving dozens of community treatment programs across the United States [50]. Interventions were prioritized for testing in CTN trials based on their established efficacy in prior research. Consequently MI and MET were chosen for testing in four multisite trials within the CTN conducted in outpatient clinics [51,52], with Spanish-speakers [53], and with pregnant drug users [54]. Research methodology was consistently strong, and in all four trials, no significant main effect of MI or MET (relative to treatment as usual) was found on the specified endpoint measures. In two of the trials [51,53] intervention by time interactions reflected a significant advantage for MET during the follow-up period after treatment completion, a period often regarded to be irrelevant in pharmaceutical trials [55]. Two trials also reported site by treatment interactions such that the effectiveness of the treatment varied significantly across sites [51,54].

Other multisite trials have tested MI-based interventions compared with no intervention [56], a waiting list [57], treatment as usual [58,59], or more extensive treatment [46,57,60–63]. Again effects have varied across trials, and the effectiveness of intervention has differed from site to site within multisite studies [61,63]. Although it may seem odd for a treatment to work at one site and not another, site-by-treatment interactions are common (though not always reported) in multisite trials, also occurring in placebo-controlled medication trials [64].

A possible source of site and study differences is that client outcomes vary widely across the therapists delivering treatments [25,63,65]. Providers are nested within and thereby confounded with sites and studies. Interventionists vary in their level and consistency of treatment fidelity, but may also differ in other outcome-relevant ways that are not detected by QA measures. Only a portion of therapist effects is accounted for by MI fidelity measures.

These multisite trials point to three potential factors that deserve closer study in MI research as well as for other behavioral interventions. First, MI "works" at some sites and not others, suggesting unidentified contextual factors influencing the efficacy of MI. Rather than being regarded as unwanted noise, site-by-treatment interactions may yield important information. Studies may identify attributes of implementation procedures, providers, client populations, or treatment settings that influence effectiveness. Second, therapists vary widely in MI skill and client outcomes, which may account for observed efficacy differences across sites and studies. Finally, benefits of a behavioral intervention may emerge only after the active treatment phase, so length of follow-up and choice of endpoints are important considerations [55].

7. Effectiveness studies

Compared to efficacy trials, effectiveness studies are meant to assess the impact of interventions when delivered "in the real world," typically in ongoing community clinics in the hands of regular agency staff. With pharmacotherapies such studies are reasonably straightforward. Medications are delivered "open label" without double-blind and sometimes without randomization, as would occur in normal practice. The question is whether the positive medication effects observed in highly-controlled efficacy trials obtain under conditions of ordinary treatment delivery.

What then constitutes an adequate effectiveness trial of a complex behavioral intervention? One perspective is that extensive training time and ongoing QA are not feasible in ordinary practice, and therefore a realistic test of any intervention is whether it can be achieved within the constraints of normal service delivery (e.g., with minimal training). Though understandable, this is analogous to testing the effectiveness of a medication capsule with unknown and highly variable contents. Without QA measures (commonly absent in normal practice) it is impossible to know what was "in the capsule," what was actually being delivered and how well. Without adequate QA, these are essentially trials of a dose of training and not of the particular intervention itself. For example, in a large effectiveness trial of MI delivered by nurses via recorded telephone conversations, subsequent process analyses revealed that key aspects of MI were rarely present in the sessions [66]. Even in controlled trials, off-protocol "chat" is a frequent element and has been related to poorer client outcomes [67].

A legitimate effectiveness question, we believe, is whether it is cost-effective to train (or re-train) clinical staff to a level of competency in a new intervention. If staff are properly trained to provide the intervention and do deliver it with fidelity, are client outcomes significantly improved over those resulting from prior treatment procedures? Such analyses can take into account the additional time and cost investment required in training and providing the new intervention [68,69]. Is the additional cost offset by improved outcomes?

8. Treatment process research

Finding an effective intervention is only part of a puzzle. A remaining question is how the intervention exerts its effects. It has been common in medicine to discover, document and deliver interventions that have beneficial effects for unknown reasons. The health benefits of carrying citrus fruit on sailing ships [70] or of washing one's hands before delivering a baby [71] were documented long before the reasons were known. Placebo effects can be controlled for, but the mechanisms by which a particular medication works may be understood poorly or not at all. Subsequent clarification of the mechanisms of efficacy can lead to new discoveries and still better treatments.

Increasing attention has been paid during the past two decades to exploring mechanisms of action in behavioral interventions [72–74]. As "active ingredients" of treatment are identified (including those sometimes described as non-specific, interpersonal, or general factors) it becomes clearer what should be emphasized in implementation, training, and QA.

Thus far Miller and Rollnick [2,75] have identified three mechanisms of change within MI for which there is reasonable empirical support. First is the therapeutic skill of accurate empathy as described by Rogers and his students [21,22,76]. Empathy is sometimes treated as a non-specific factor in psychotherapy [77], but within MI it is a fundamental, purposeful, specific and measurable component skill. Within addiction treatment, therapist empathy is rather consistently associated with better client outcomes, and low empathy with poor outcomes [78].

Therapist empathy in MI has also been linked to change talk [79], better outcomes [80], and lower client resistance [81]. Secondly, training in MI may suppress counter-therapeutic responses such as confrontation and counterargument that are linked to increased client resistance and poorer outcomes [82–87]. Thirdly, MI training and fidelity have been linked to increased client change talk and decreased sustain talk, which in turn predict behavior change [40–42,80,84,88–91]. A meta-analysis [92] similarly identified three aspects of MI associated with better outcomes: client experience of discrepancy, low levels of MI-inconsistent counselor behavior, and client change talk.

9. Recommendations for future evaluation research

From the foregoing discussion we offer here a set of recommendations for future MI clinical research and training. Many of these apply as well to research with complex behavioral interventions more generally.

9.1. Intervention content

As knowledge emerges regarding active mechanisms at work within the practice of MI, future training and research should ensure that these are reflected in intervention content, QA, fidelity coding, and process analyses. At present we suggest that any MI intervention should contain at least these three elements reflecting the engaging, focusing, and evoking processes of MI [2]:

1. Proficiency in a person-centered counseling style and specifically in the therapeutic skill of accurate empathy
2. Clear identification of one or more change goals toward which the intervention is directed
3. Differential evocation of clients' own motivational statements (change talk) for and commitment to change.

Furthermore, an intervention identified as MI should not contain standard elements that are inconsistent with the spirit and style of MI such as confrontation or uninvited advice. We also regard a classic decisional balance [93] to be different from and inconsistent with MI, both because it is theoretically contraindicated [2,17] and because clinical trials have found that a decisional balance intervention with ambivalent people decreases their commitment to change [18]. If a therapist manual is used, it should allow substantial clinical flexibility to respond to clients' behavior in the moment, avoiding rigid prescription of required steps, sequence, or procedures (such as a change plan) without regard to client response.

9.2. Training

Providers in clinical studies should be trained to a specified within-trial criterion of proficiency based on observed practice. No fixed dose of training is sufficient for all; the criterion is the ability to deliver MI at a specified standard of competence. Provisional thresholds for competent practice have been suggested [2], though these are likely to require adjustment to particular contexts. The key is to have a specified criterion performance level that all providers must demonstrate before being certified and beginning to deliver MI in evaluation research. Studies should document and report pre-trial levels of proficiency using reliable practice coding procedures.

9.3. Quality assurance and fidelity

Pre-trial competence to deliver MI does not guarantee fidelity in doing so during the study interventions [16]. We recommend routine recording of all sessions, and for most purposes audio recording is sufficient. This permits reporting of actual intervention fidelity during the study period. Use a replicable coding system, with training of coders to a specified criterion of inter-rater reliability [37,94]. If QA is not done for all sessions, select a random sample for coding, with providers unaware of which sessions will be coded. In efficacy trials the QA process should be rapid and ongoing so that drift in practice quality can be detected and corrected quickly throughout the study. To maintain a standard of intervention quality, give providers corrective feedback in timely fashion, with close fidelity monitoring of subsequent sessions. If intervention quality falls below a specified standard, a provider can be decertified from treating further study cases until criterion proficiency has been restored [95].

High standards for training and fidelity monitoring should be maintained in effectiveness as well as efficacy studies. In differentiating explanatory (efficacy) from pragmatic (effectiveness) trials, Thorpe and colleagues [96] suggested that in the latter an intervention should be delivered by providers “regardless of their expertise” and that there should be “unobtrusive (or no) measurement of compliance” (p.466). Although this logic can apply to pharmacologic trials where medication content is known and standardized, it does not generalize well to complex behavioral interventions where specific provider expertise is required and one cannot know whether the intervention is being tested without knowing if it was delivered. Even with previously demonstrated efficacy, the effectiveness of an intervention in standard practice cannot be determined if the intervention has not been provided.

In the case of MI, a sharp explanatory/pragmatic distinction is problematic. Thorpe et al. [96]. characterized explanatory trials as having an “inflexible experimental intervention, with strict instructions for every element” (p. 466). Such inflexibility is incompatible with MI, in which the provider's behavior is guided by clients' immediate responses, and (as described earlier) prescribed sequencing would be expected to undermine efficacy. In pragmatic trials, after provider proficiency has been established, a reasonable compromise to parallel standard practice conditions is to monitor fidelity without offering ongoing corrective feedback [96].

9.4. Process coding

QA systems like the MITI that code only therapist responses offer a partial picture of MI. If evoking client change talk is one key mechanism of action in MI, this can be measured only by coding client responses as well. Certain practice behaviors are associated with increased frequency of client change talk [40], but do not guarantee its occurrence. Coding both provider and client responses enables the linkage of in-session communication processes to treatment outcomes [38,89].

10. Summary and conclusions

One possible explanation of negative trials of a complex behavioral intervention is that there is no effect to detect: that the treatment is inert and exerts no specific effect. Most meta-analyses of MI have reached a different conclusion, but a clear pattern in efficacy trials of interventions identified with MI is high variability in outcomes across studies, sites, and therapists. Effects are of small to medium size on average, ranging from nil to quite large effects. It is important to clarify the conditions under which a complex treatment like MI is less or more effective. Larger effects have been linked to a number of factors including absence of a therapist manual [4], intervention fidelity [28–35], and disadvantaged minority populations [4,54,97].

We regard an adequate efficacy or effectiveness study of MI to be one in which at least the three above-described fidelity conditions are met. Of the many published outcome studies of interventions identified as MI, a relatively small subset would meet this methodological standard for demonstrating that MI was actually delivered, and this subset contains both positive and negative findings. These considerations are not unique to MI. In any study of a complex behavioral intervention there are legitimate concerns as to whether the treatment contained an adequate dose of the crucial active ingredients, the providers were trained to an appropriate level of competence, and the intervention was actually provided with fidelity sufficient to expect an effect.

This suggests a different approach in evaluating the efficacy and effectiveness of complex behavioral treatments. Whereas meta-analyses to date have typically included all studies delivering interventions that were identified as MI, a more refined approach would be limited to studies in which there is clear evidence that MI was actually provided. Most likely there are studies in which fidelity was not documented but MI was nevertheless delivered; it is just impossible to tell without QA monitoring and reporting. This involves more than including the three fidelity conditions as moderator variables in a meta-analysis, because the review would then include (and be biased by) studies in which MI was never delivered. (An analogy could be a meta-analysis of the efficacy of medications in which contents of the administered capsule were examined as moderators of effect.)

Another useful analytic approach would be dose–response analyses relating behavioral outcomes to the amount and quality of exposure to MI. There is already evidence for a dose effect in meta-analyses of MI trials [6,98]. Instead of conceptualizing MI as monolithic, it could also be useful to evaluate the importance of fidelity to various components of MI (such as empathy [99], direction [100], and the elicitation of change talk [88–90]). Such analyses could help to clarify the effective processes at work within a complex behavioral intervention like MI.

Finally we note that different behavioral treatment components are often blended to create a hybrid intervention. There have been numerous trials in which MI was combined with standard practice or other evidence-based interventions. If such studies do not meet the fidelity criteria for MI described in this article, it does not invalidate them as evaluations of the particular intervention that was delivered. A problem arises, however, if such studies are subsequently represented (without sufficient evidence) as evaluations of MI itself. Optimal procedures for training, QA, and interpretation of such hybrid interventions remain to be determined.

References

[1] Fixsen DL, Naoom SF, Blase KA, Friedman RM, Wallace F. Implementation research: a synthesis of the literature. Tampa, FL: University of South Florida, National Implementation Research Network; 2005.
[2] Miller WR, Rollnick S. Motivational interviewing: helping people change. 3rd ed. New York: Guilford Press; 2013.
[3] McLouth CJ. An expanded bibliography of motivational interviewing. http://www.guilford.com/add/miller2/biblio.pdf; 2013.
[4] Hettema J, Steele J, Miller WR. Motivational interviewing. Annu Rev Clin Psychol 2005;1:91–111.
[5] Lundahl B, Burke BL. The effectiveness and applicability of motivational interviewing: a practice-friendly review of four meta-analyses. J Clin Psychol 2009;65:1232–45.
[6] Lundahl BW, Kunz C, Brownell C, Tollefson D, Burke BL. A meta-analysis of motivational interviewing: twenty-five years of empirical studies. Res Soc Work Pract 2010;20(2):137–60.
[7] Vasilaki EI, Hosier SG, Cox WM. The efficacy of motivational interviewing as a brief intervention for excessive drinking: a meta-analytic review. Alcohol Alcohol 2006;41(3):328–35.
[8] Tait RJ, Hulse GK. A systematic review of the effectiveness of brief interventions with substance using adolescents by type of drug. Drug Alcohol Rev 2003;22:337–46.
[9] Smedslund G, Berg RC, Hammerstrøm KT, et al. Motivational interviewing for substance abuse. Cochrane Database Syst Rev 2011;5:CD008063.
[10] Grimshaw G, Stanton A. Tobacco cessation interventions for young people. Cochrane Database Syst Rev 2010;4:CD003289.
[11] Armstrong MJ, Mottershead RA, Ronksley PE, Sigal RJ, Campbell TS, Hemmelgarn BR. Motivational interviewing to improve weight loss in overweight and/or obese patients: a systematic review and meta-analysis of randomized clinical trials. Obes Rev 2011;12(9):709–23.
[12] Gooding P, Tarrier N. A systematic review and meta-analysis of cognitive-behavioural interventions to reduce problem gambling: hedging our bets? Behav Res Ther 2009;47(4):592–607.
[13] Lundahl B, Moleni T, Burke BL, et al. Motivational interviewing in medical care settings: a systematic review and meta-analysis of randomized controlled trials. Patient Educ Couns 2013;93(2):157–68.
[14] Miller WR, Yahne CE, Tonigan JS. Motivational interviewing in drug abuse services: a randomized trial. J Consult Clin Psychol 2003;71:754–63.
[15] Robling M, McNamara R, Bennert K, et al. The effect of the Talking Diabetes consulting skills intervention on gycaemic control and quality of life in children with type 1 diabetes: cluster randomised controlled trial (DEPICTED study). Br Med J 2012;344.
[16] Bellg A, Borelli B, REsnick B, et al. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the Behavioral Change Consortium. Health Psychol 2004;23(5):443–51.
[17] Miller WR, Rollnick S. Ten things that motivational interviewing is not. Behav Cogn Psychother 2009;37:129–40.
[18] Miller WR, Rose GS. Motivational interviewing and decisional balance: contrasting responses to client ambivalence. Behav Cogn Psychother 2013. http://dx.doi.org/10.1017/S1352465813000878 [in press].
[19] Kuchipudi V, Hobein K, Fleckinger A, Iber FL. Failure of a 2-hour motivational intervention to alter recurrent drinking behavior in alcoholics with gastrointestinal disease. J Stud Alcohol 1990;51:356–60.
[20] Björk A. Stabilizing a fluid intervention: the development of motivational interviewing 1983–2013. Addict Res Theory 2013. http://dx.doi.org/10.3109/16066359.2013.845174 [in press].
[21] Rogers CR. Client-centered therapy. New York: Houghton-Mifflin; 1965.
[22] Truax CB, Carkhuff RR. Toward effective counseling and psychotherapy. Chicago: Aldine; 1967.
[23] Meyers RJ, Miller WR. A community reinforcement approach to addiction treatment. Cambridge, UK: Cambridge University Press; 2001.
[24] Rollnick S. Comments on Dunn et al's "The use of brief interventions adapted from motivational interviewing across behavioral domains: a systematic review". Enthusiasm, quick fixes and premature controlled trials. Addiction 2001;96(12):1769–70.
[25] Project MATCH Research Group. Therapist effects in three treatments for alcohol problems. Psychother Res 1998;8:455–74.
[26] Tober G, Clyne W, Finnegan O, A. F., Ian Russell Team, UKATT Research Team. Validation of a scale for rating the delivery of psycho-social

treatments for alcohol dependence and misuse: the UKATT process rating scale (PRS). Alcohol Alcohol 2008;43(6):675–82.
[27] Madson MB, Campbell TC. Measures of fidelity in motivational enhancement: a systematic review. J Subst Abus Treat 2006;31:67–73.
[28] Daeppen JB, Bertholet N, Gaume J. What process research tells us about brief intervention efficacy. Drug Alcohol Rev 2010;29:612–6.
[29] Gaume J, Gmel G, Faouzi M, Daeppen JB. Counselor skill influences outcomes of brief motivational interventions. J Subst Abus Treat 2009;37(2):151–9.
[30] Magill M, Mastroleo NR, Apodaca TR, Barnett NP, Colby SM, Monti PM. Motivational interviewing with significant other participation: assessing therapeutic alliance and patient satisfaction and engagement. J Subst Abus Treat 2010;39(4):391–8.
[31] Moyers TB, Miller WR, Hendrickson SML. How does motivational interviewing work? Therapist interpersonal skill predicts client involvement within motivational interviewing sessions. J Consult Clin Psychol 2005;73:590–8.
[32] Pollak KI, Alexander SC, Coffman C, et al. Physician communication techniques and weight loss in adults: project CHAT. Am J Prev Med 2010;39(4):321–8.
[33] Vader AM, Walters ST, RG C, Houck JM, Field CA. The language of motivational interviewing and feedback: counselor language, client language, and client drinking outcomes. Psychol Addict Behav 2010;24(2):190–7.
[34] Smith DC, Hall JA, Jang M, Arndt S. Therapist adherence to a motivational interviewing intervention improves treatment entry for substance-misusing adolescents with low problem perception. J Stud Alcohol Drugs 2009;70(1):101–5.
[35] Pollak KI, Alexander SC, Østbye T, et al. Primary care physicians' discussions of weight-related topics with overweight and obese adolescents: results from the Teen CHAT pilot study. J Adolesc Health 2009;45(2):205–7.
[36] Moyers TB, Martin T, Manuel JK, Hendrickson SML, Miller WR. Assessing competence in the use of motivational interviewing. J Subst Abus Treat 2005;28:19–26.
[37] Pierson HM, Hayes SC, Gifford EV, et al. An examination of the Motivational Interviewing Treatment Integrity code. J Subst Abus Treat 2007;32:11–7.
[38] Moyers TB, Martin T, Catley D, Harris KJ, Ahluwalia JS. Assessing the integrity of motivational interventions: reliability of the Motivational Interviewing Skills Code. Behav Cogn Psychother 2003;31:177–84.
[39] DeJonge JJM, Schippers GM, Schaap CPDR. The Motivational Interviewing Skill Code: reliability and a critical appraisal. Behav Cogn Psychother 2005;33:1–14.
[40] Moyers TB, Martin T. Therapist influence on client language during motivational interviewing sessions. J Subst Abus Treat 2006;30:245–52.
[41] Moyers TB, Martin T, Christopher PJ, Houck JM, Tonigan JS, Amrhein PC. Client language as a mediator of motivational interviewing efficacy: where is the evidence? Alcohol Clin Exp Res 2007;31:40S–7S [Supplement].
[42] Moyers TB, Martin T, Houck JM, Christopher PJ, Tonigan JS. From in-session behaviors to drinking outcomes: a causal chain for motivational interviewing. J Consult Clin Psychol 2009;77(6):1113–24.
[43] Miller WR, Sorensen JL, Selzer J, Brigham G. Disseminating evidence-based practices in substance abuse treatment: a review with suggestions. J Subst Abus Treat 2006;31:25–39.
[44] Miller WR, Yahne CE, Moyers TB, Martinez J, Pirritano M. A randomized trial of methods to help clinicians learn motivational interviewing. J Consult Clin Psychol 2004;72:1050–62.
[45] Broekhuizen K, Jelsma JGM, van Poppel MNM, Koppes LLJ, Brug J, van Mechelen W. Is the process of delivery of an individually tailored lifestyle intervention associated with improvements in LDL cholesterol and multiple lifestyle behaviours in people with familial hypercholesterolemia? BMC Public Health 2012;12(348).
[46] Rapp RC, Otto AL, Lane DT, Redko C, McGatha S, Carlson RG. Improving linkage with substance abuse treatment using brief case management and motivational interviewing. Drug Alcohol Depend 2008;94(1–3):172–82.
[47] Miller WR. Form 90: a structured assessment interview for drinking and related behaviors. , Vol. 5Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 1996.
[48] Westerberg VS, Tonigan JS, Miller WR. Reliability of Form 90D: an instrument for quantifying drug use. Subst Abus 1998;19:179–89.
[49] Amrhein PC, Miller WR, Yahne CE, Palmer M, Fulcher L. Client commitment language during motivational interviewing predicts drug use outcomes. J Consult Clin Psychol 2003;71:862–78.
[50] Tai B, Straus MM, Liu DS, Sparenborg S, JT R, McCarty D. The first decade of the National Drug Abuse Treatment Clinical Trials Network: bridging the gap between research and practice to improve drug abuse treatment. J Subst Abus Treat 2010;38(Suppl. 1):S4–S13.
[51] Ball SA, Martino S, Nich C, et al. Site matters: multisite randomized trial of motivational enhancement therapy in community drug abuse clinics. J Consult Clin Psychol 2007;75:556–67.
[52] Carroll KM, Ball SA, Nich C, et al. Motivational interviewing to improve treatment engagement and outcome in individuals seeking treatment for substance abuse: a multisite effectiveness study. Drug Alcohol Depend 2006;81:301–12.
[53] Carroll KM, Martino S, Ball SA, et al. A multisite randomized effectiveness trial of motivational enhancement therapy for Spanish-speaking substance users. J Consult Clin Psychol 2009;77(5):993–9.
[54] Winhusen T, Kropp F, Babcock D, et al. Motivational enhancement therapy to improve treatment utilization and outcome in pregnant substance users. J Subst Abus Treat 2008;35:161–73.
[55] Miller WR, LoCastro JS, Longabaugh R, O'Malley S, Zweben A. When worlds collide: blending the divergent traditions of pharmacotherapy and psychotherapy outcome research. J Stud Alcohol 2005;15:17–23.
[56] McCambridge J, Strang J. The efficacy of single-session motivational interviewing in reducing drug consumption and perceptions of drug-related risk and harm among young people: results from a multi-site cluster randomized trial. Addiction 2004;99:39–52.
[57] Marijuana Treatment Project Research Group. Brief treatments for cannabis dependence: findings from a randomized multisite trial. J Consult Clin Psychol 2004;72:455–66.
[58] Heisler M, Hofer TP, Schmittdiel JA, et al. Improving blood pressure control through a clinical pharmacist outreach program in patients with diabetes mellitus in 2 high-performing health systems: the adherence and intensification of medications cluster randomized, controlled pragmatic trial. Circulation 2012;125:2863–72.
[59] Naar-King S, Parsons JT, Murphy D, Kolmodin K, Harris DR. A multisite randomized trial of a motivational intervention targeting multiple risks in youth living with HIV: initial effects on motivation, self-efficacy, and depression. J Adolesc Health 2010;46(5):422–8.
[60] Koblin B, Chesney M, Coates T, Team ES. Effects of a behavioural intervention to reduce acquisition of HIV infection among men who have sex with men: the EXPLORE randomised controlled study. Lancet 2004;364(9428):41–50.
[61] Project MATCH Research Group. Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcomes. J Stud Alcohol 1997;58:7–29.
[62] UKATT Research Team. Effectiveness of treatment for alcohol problems: findings of the randomized UK alcohol treatment trial (UKATT). Br Med J 2005;331:541–4.
[63] Madras BK, Compton WM, Avula D, Stegbauer T, Stein JB, Clark HW. Screening, brief interventions, referral to treatment (SBIRT) for illicit drug and alcohol use at multiple healthcare sites: comparison at intake and 6 months later. Drug Alcohol Depend 2009;99:280–95.
[64] Anton RF, O'Malley SS, Ciraulo DA, et al. Combined pharmacotherapies and behavioral interventions for alcohol dependence: the COMBINE study: a randomized controlled trial. JAMA 2006;295(17):2003–17.
[65] Copello A, Godfrey C, Heather N, et al. United Kingdom Alcohol Treatment Trial (UKATT): hypotheses, design and methods. Alcohol Alcohol Jan–Feb 2001;36:11–21.
[66] Ernst DB. Motivational interviewing and health coaching: a quantitative and qualitative exploration of integration. Albuquerque, NM: Department of Psychology, The University of New Mexico; 2008 [Dissertation].
[67] Bamatter W, Carroll KM, Añez LM, et al. Informal discussions in substance abuse treatment sessions with Spanish-speaking clients. J Subst Abus Treat 2010;39(4):353–63.
[68] Cisler R, Holder HD, Longabaugh R, Stout RL, Zweben A. Actual and estimated replication costs for alcohol treatment modalities: case study from Project MATCH. J Stud Alcohol Sep 1998;59(5):503–12.
[69] Holder HD, Cisler RA, Longabaugh R, Stout RL, Treno AJ, Zweben A. Alcoholism treatment and medical care costs from Project MATCH. Addiction 2000;95:999–1013.
[70] Rogers EM. Diffusion of innovations. 5th ed. New York: Free Press; 2003.
[71] Broad W, Wade N. Betrayers of the truth: fraud and deceit in the halls of science. New York: Simon & Schuster; 1982.
[72] Morgenstern J, Blanchard KA, Kahler C, Barbosa KM, McCrady BS, McVeigh KH. Testing mechanisms of action for intensive case management. Addiction Mar 2008;103(3):469–77.
[73] Morgenstern J, Longabaugh R. Cognitive–behavioral treatment for alcohol dependence: a review of evidence for its hypothesized mechanisms of action. Addiction 2000;95:1475–90.
[74] Project MATCH hypotheses: results and causal chain analyses. In: Longabaugh R, Wirtz PW, editors. Project MATCH Monograph

Series; No. Project MATCH Monograph Series, Vol. 8. Bethesda, Maryland: National Institute on Alcohol Abuse and Alcoholism; 2001.
[75] Miller WR, Rose GS. Toward a theory of motivational interviewing. Am Psychol 2009;64:527–37.
[76] Rogers CR. A theory of therapy, personality, and interpersonal relationships as developed in the client-centered framework. In: Koch S, editor. Psychology: the study of a science. , Formulations of the person and the social contextsNew York: McGraw-Hill; 1959. p. 184–256.
[77] Bohart AC, Elliott R, Greenberg LS, Watson JC. Empathy. In: Norcross JC, editor. Psychotherapy relationships that work. New York: Oxford University Press; 2002. p. 89–108.
[78] Moyers TB, Miller WR. Is low therapist empathy toxic? Psychol Addict Behav 2013;27(3):878–84.
[79] Fischer DJ, Moyers TB. Is there an association between empathic speech and change talk in motivational interviewing sessions? Alcohol Treat Q 2014;32(1):3–18.
[80] Kaplan JE, Keeley RD, Engel M, Emsermann C, Brody D. Aspects of patient and clinician language predict adherence to antidepressant medication. J Am Board Fam Med 2013;26:409–20.
[81] Grodin JP. Dissertation , Assessing therapeutic change mechanisms in motivational interviewing using the articulated thoughts in simulated situations paradigm. Los Angeles: Psychology, University of Southern California; 2006.
[82] Baer JS, Carpenter KM, Beadnell B, et al. Computer assessment of simulated patient interviews (CASPI): psychometric properties of a web-based system for the assessment of motivational interviewing skills. J Stud Alcohol Drugs 2012;73(1):154–64.
[83] Miller WR, Mount KA. A small study of training in motivational interviewing: does one workshop change clinician and client behavior? Behav Cogn Psychother 2001;29:457–71.
[84] Glynn LH, Moyers TB. Chasing change talk: the clinician's role in evoking client language about change. J Subst Abus Treat 2010;39:65–70.
[85] Miller WR, Benefield RG, Tonigan JS. Enhancing motivation for change in problem drinking: a controlled comparison of two therapist styles. J Consult Clin Psychol 1993;61:455–61.
[86] Patterson GR, Forgatch MS. Therapist behavior as a determinant for client noncompliance: a paradox for the behavior modifier. J Consult Clin Psychol 1985;53:846–51.
[87] White WL, Miller WR. The use of confrontation in addiction treatment: history, science, and time for change. Counselor 2007;8(4):12–30.
[88] Bertholet N, Faouzi M, Gmel G, Gaume J, Daeppen JB. Change talk sequence during brief motivational intervention, towards or away from drinking. Addiction 2010;105:2016–112.
[89] Daeppen J-B, Bertholet N, Gmel G, Gaume J. Communication during brief intervention, intention to change, and outcome. Subst Abus 2007;28(3):43–51.
[90] Gaume J, Bertholet N, Faouzi M, Gmel G, Daeppen JB. Counselor motivational interviewing skills and young adult change talk articulation during brief motivational interventions. J Subst Abus Treat 2010;39:272–81.
[91] Miller WR, Rollnick S. Talking oneself into change: motivational interviewing, stages of change, and therapeutic process. J Cogn Psychother 2005;18:299–308.
[92] Apodaca TR, Longabaugh R. Mechanisms of change in motivational interviewing: a review and preliminary evaluation of the evidence. Addiction 2009;104:705–15.
[93] Janis IL, Mann L. Decision making: a psychological analysis of conflict, choice and commitment. New York: Free Press; 1977.
[94] Hendrickson SML, Martin T, Manuel JK, Christopher PJ, Thiedeman T, Moyers TB. Assessing reliability of the motivational interviewing treatment integrity behavioral coding system under limited range. Alcohol Clin Exp Res May 2004;28(5):74A.
[95] Miller WR, Moyers TB, Arciniega LT, Ernst D, Forcehimes A. Training, supervision and quality monitoring of the COMBINE study behavioral interventions. J Stud Alcohol 2005;15:188–95.
[96] Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J Clin Epidemiol 2009;62:464–75.
[97] Villanueva M, Tonigan JS, Miller WR. Response of Native American clients to three treatment methods for alcohol dependence. J Ethn Subst Abus 2007;6(2):41–8.
[98] Burke BL, Arkowitz H, Menchola M. The efficacy of motivational interviewing: a meta-analysis of controlled clinical trials. J Consult Clin Psychol 2003;71:843–61.
[99] Miller WR, Taylor CA, West JC. Focused versus broad spectrum behavior therapy for problem drinkers. J Consult Clin Psychol 1980;48:590–601.
[100] Karno MP, Longabaugh R. An examination of how therapist directiveness interacts with patient anger and reactance to predict alcohol use. J Stud Alcohol 2005;66:825–32.