Contemporary Clinical Trials 43 (2015) 93–99

Contents lists available at ScienceDirect

Contemporary Clinical Trials

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

How to Measure Motivational Interviewing Fidelity in Randomized Controlled Trials: Practical Recommendations

Judith G.M. Jelsma [a,*,1], Vera-Christina Mertens [b,1], Lisa Forsberg [c], Lars Forsberg [c]

[a] Department of Public and Occupational Health, EMGO+ Institute for Health and Care Research, VU University Medical Center Amsterdam, The Netherlands
[b] Department of Rehabilitation Medicine, CAPHRI School for Public Health and Primary Care, Maastricht University, The Netherlands
[c] Department of Clinical Neuroscience, MIC Lab, Karolinska Institutet, Stockholm, Sweden

ARTICLE INFO

Article history:
Received 22 January 2015
Received in revised form 30 April 2015
Accepted 2 May 2015
Available online 9 May 2015

Keywords:
Motivational interviewing
Fidelity
Randomized controlled trials
Motivational Interviewing Treatment Integrity Code
Motivational interviewing/methods*

ABSTRACT

Many randomized controlled trials in which motivational interviewing (MI) is a key intervention make no provision for the assessment of treatment fidelity. This methodological shortcoming makes it impossible to distinguish between high- and low-quality MI interventions, and, consequently, to know whether MI provision has contributed to any intervention effects. This article makes some practical recommendations for the collection, selection, coding and reporting of MI fidelity data, as measured using the Motivational Interviewing Treatment Integrity Code. We hope that researchers will consider these recommendations and include MI fidelity measures in future studies.

1. Introduction

"Motivational interviewing (MI) is a collaborative, goal-oriented style of communication with particular attention to the language of change. It is designed to strengthen personal motivation for and commitment to a specific goal by eliciting and exploring the person's own reasons for change within an atmosphere of acceptance and compassion" [1]. MI has been shown to be superior to no intervention, and as efficacious as other evidence-based interventions in systematic reviews and meta-analyses across a variety of different problem behaviors and health care settings [2–6].

A randomized controlled trial (RCT) tests whether an intervention is efficacious in an ideal situation by maximizing internal validity through controlling all variables except the intervention to be tested. A controlled clinical trial tests instead whether an intervention is effective in a real life setting, maximizing the external validity to ensure generalizability [7]. In order for the move from an efficacy trial to an effectiveness trial to be successful, it is important to have identified the active mechanism of the intervention tested in the efficacy study [8]. In respect of behavior change intervention research, the reporting of treatment fidelity is likely to improve the credibility of evidence that results from a trial [9]. Treatment fidelity refers to the "methodological strategies used to monitor and enhance the reliability and validity of behavioral interventions" [9] (p. 443).

Treatment fidelity in MI has predictive validity in respect of patient behavior following the intervention [10–12]. However, many research trials conducted have failed to assess treatment fidelity of the intervention that is being delivered. This makes it impossible to ascertain whether the result can accurately be attributed to the MI intervention, that is, whether we can in fact be sure that MI is the actual working mechanism that is "doing the job" [13]. Miller and Rollnick (2014) suggest that treatment fidelity should be assessed throughout a study, through a reliable assessment procedure ('coding'), and be reported in a manner that allows for comparison across trials [13].

The aim of this paper is to provide guidance to researchers in respect of assessing and reporting MI treatment fidelity. The practical recommendations offered are important to consider in designing, developing and conducting research, including in grant applications.

2. The Motivational Interviewing Treatment Integrity (MITI) Code

The Motivational Interviewing Treatment Integrity (MITI) code is the most frequently used [13] instrument for assessing MI fidelity in

* Corresponding author at: Department of Public and Occupational Health, EMGO+ Institute for Health and Care Research, VU University Medical Center Amsterdam, Van der Boechorststraat 7, 1081BT Amsterdam, The Netherlands.
E-mail address: j.jelsma@vumc.nl (J.G.M. Jelsma).
[1] Authors contributed equally to this work.

http://dx.doi.org/10.1016/j.cct.2015.05.001
1551-7144/

RCTs [14–16]. The MITI has been derived from the Motivational Interviewing Skill Code (MISC) [17], and while reducing the MISC's complexity and length [18], the MITI focuses exclusively on the verbal behavior of the practitioner, and does not take client responses into account [15]. The MITI is continuously revised and improved. Almost ten years ago, MITI version 2.0 was being used, and at present, the MITI 4.1 has just been published. Definitions of variables that measure important aspects of MI practice are improved in each new version, with the aim to carefully follow and progress developments in MI research. Every previous version of the MITI instrument [14] has been shown to be reliable [15,16,19,20] and valid [14,16]. The recently published MITI 4.1 has been shown to have face validity, but the collection of data regarding its validity and reliability is still underway.

In the recent MITI 4.1, substantial changes have been made in comparison to MITI 3.1.1. The authors claim that the two versions are not comparable, and advise researchers to use the MITI 4.1 from now on. However, before MITI 4.1 may be used more widely, the instrument has to show predictive validity at least in respect of some problem behaviors and in some languages, and coders using the new instrument need to be able to achieve an adequate inter-rater reliability. In order to assist researchers both in conducting treatment fidelity assessment in future research, and in interpreting research conducted so far, both versions of the MITI are briefly discussed in this paper.

The MITI comprises two separate components: global variables and behavior counts. In both versions of the MITI, a 20-minute segment is used both for the behavior counts and for rating the global variables. The global ratings reflect the coder's overall impression of how well or poorly a practitioner performed in a certain aspect of MI practice, rated on a five point Likert scale. In MITI 3.1.1, the global scales are Empathy, Evocation, Collaboration, Autonomy/Support, and Direction. In both MITI 3.1.1 and MITI 4.1, the Empathy rating captures how well the practitioner understands the client's perspective, experiences, and feelings. In MITI 3.1.1, the global variable MI Spirit combines the ratings of Evocation, Collaboration and Autonomy/Support (by taking an average of the ratings of all three variables). In MITI 4.1, the MI Spirit variable is replaced by the variables Cultivating Change Talk, Softening Sustain Talk, and Partnership. This modification emphasizes the importance of the practitioner adapting her behavior in response to client utterances. A further change in respect of the global variables in MITI 4.1 is that the Direction variable has been removed.

The behavior counts are intended to capture specific practitioner verbal behaviors that are relevant to good practice of MI. The MITI 4.1 retains the behavior count categories Giving Information, Simple Reflections, and Complex Reflections. However, some other behavior count categories have been changed in the MITI 4.1. First, the two categories Open and Closed Questions have been combined into one single Questions category. Second, the category for MI Adherent practitioner behavior has been divided into several categories for subtypes of such behavior, each of which is given a separate code: Seeking Collaboration, Emphasizing Autonomy, and Affirm. Third, the category for MI Non-Adherent practitioner behavior in MITI 3.1.1 has been split up into Confront, and Persuade (with and without permission) in MITI 4.1. For a more detailed discussion of the different variables in the two different versions of the MITI we refer to the MITI 3.1.1 manual [14] and the MITI 4.1 manual [21].

3. Assessing Treatment Fidelity in MI Sessions

Prior to the start of an RCT where MI is one of the interventions being tested, it is important to consider the following three things: 1) which samples of MI practice (sessions) will be collected and selected for fidelity assessment; 2) who will do the assessment (coding) of these sessions; and 3) how will the results be reported. These three essential questions will be discussed in detail below.

3.1. Collection of Audio-recorded Sessions

Since audio-recorded sessions are used to assess treatment fidelity — that is, what really happened in the interaction between practitioner and client — it is important to audio record all, or as much as possible, of the conducted conversations. Recording all sessions allows the researcher to minimize selection bias, which is easily introduced if practitioners are permitted to select the sessions submitted for treatment integrity assessment themselves [22]. Approval by the relevant ethics review board, and the consent of clients and practitioners, have to be obtained prior to the audio recording. Informing clients and practitioners that the data will be anonymized might make them less reluctant to consent. In addition, providing practitioners with digital audio recording devices (and checking compliance throughout the study) could assist in obtaining the full spectrum of conversations.

3.2. Selection of Samples for Assessment

A random representative sample of the collected audio-recorded sessions should be selected. It will often not be possible to assess the treatment fidelity of all sessions, but coding multiple work samples from each practitioner provides a more accurate assessment of his or her proficiency [23]. So the question is, how large should this representative sample be, keeping in mind that studies have different design in respect of the number of participating practitioners, the number of sessions per client, and so forth.

In previously conducted RCTs where attempts have been made to assess treatment fidelity, between 11–32% of the total number of sessions were assessed (e.g. [24] (25%); [25] (16%); [26] (11%); [27] (28%); [28] (25%); [29] (23%); [30] (32%)). However, the study of Smith et al. (2012) is an exception to this since 100% of recorded sessions were assessed, although the total number of sessions in this study only comprised 38 [31]. In studies where the intervention was delivered by more than one practitioner, 10–17 sessions per practitioner were selected for assessment ([32], (n = 17); [26] (n = 10)) to represent a reliable overview of the quality of the individual practitioner throughout the study period. El-Mallakh et al. (2012) assessed 18 sessions (25% of total sessions) [28], and McCarthy et al. (2014) assessed 4 sessions (20% of total sessions) [33] throughout the study period (both with only one practitioner delivering the intervention), providing an indication of the MI skill fluctuation in the practitioner delivering the intervention over time.

Some studies require a comparison of overall group results (average of multiple practitioners), for example, when usual care conditions containing an attention control intervention without an MI component and an MI intervention condition are compared (e.g. [31]), or when practitioners with different backgrounds/experience are compared (e.g. [29] (n = 19)). In the study of Smith et al. (2012), one practitioner delivered both the intervention and the control arm of the study [31]. Here, it was examined if MI was more pronounced in the intervention group than in the control group by assessing 20 intervention sessions and 18 control sessions.

Audio-recorded session may vary in length between 10 min and over an hour. The MITI is used to assess a 20-minute segment of each session. It may be the case that sessions shorter in length than 20 min could not be reliably coded using the MITI [20]. For longer sessions, it may be hard to decide how to choose the segment submitted for assessment, in particular since the quality of a practitioner's MI practice might fluctuate throughout a session. This was, for example, found in the psycholinguistic study of Amrhein et al. (2003), in which there was an explicit requirement for practitioners to agree a change plan with the client at the end of