ELSEVIER

Journal of Clinical Epidemiology 62 (2009) 81–90

Journal of Clinical Epidemiology

The fidelity of treatment delivery can be assessed in treatment outcome studies: a successful illustration from behavioral medicine

M. Leeuw[a,b,*], M.E.J.B. Goossens[a], H.C.W. de Vet[c], J.W.S. Vlaeyen[a,d]

[a]*Department of Clinical Psychological Science, Faculty of Psychology, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands*

[b]*Hoensbroeck Rehabilitation Centre, Hoensbroek, The Netherlands*

[c]*EMGO Institute, VU University Medical Centre, Amsterdam, The Netherlands*

[d]*Department of Psychology, Research Centre for Health Psychology, University of Leuven, Leuven, Belgium*

Accepted 2 March 2008

Abstract

Objectives: Treatment outcome studies ought to assess the fidelity of their treatments, including treatment delivery, but practical guidelines and examples for this are lacking. Based on general recommendations in available literature, this study proposes and illustrates the design and application of a Method of Assessing Treatment Delivery (MATD) in a behavioral medicine trial comparing two treatments for chronic low back pain.

Study Design and Setting: In designing MATD, two experts identified several feasible treatment elements. Agreement between the experts in classifying these elements into five categories (essential and unique, essential but not unique, unique but not essential, compatible, prohibited) was assessed. In applying MATD, treatment recordings were evaluated by two independent raters, who coded the (non)-occurrence of MATD elements and who categorized each session as belonging to one of the two treatments.

Results: MATDs content validity was supported by adequate agreement between the experts' classifications of the treatment elements. MATDs interrater reliability was good.

Conclusion: Comprehensive illustrations of designing and applying MATD may encourage the verification of treatment delivery as a partial reflection of treatment fidelity in forthcoming treatment outcome studies.

Keywords: Process assessment (health care); Guideline adherence; Delivery of health care; Treatment outcome; Treatment fidelity; Treatment contamination

1. Introduction

Treatment fidelity, which may also be denoted as integrity, is defined as the extent to which a treatment is carried out as intended [1]. Treatment outcome studies ought to assess the fidelity of the treatments delivered to ensure honest and genuine comparisons [1–4]. Cook and Campbell [5] even asserted that ''measures of the exact nature of the treatment in *all* treatment and control groups are absolutely vital in any experiment'' (p0.59). Especially psychological interventions may be at higher risk of compromised treatment fidelity, because these are generally more complex and extensive [1–3,6]. The absence of treatment fidelity checks can seriously obscure the conclusions about treatment effectiveness: in case a treatment is found to be effective, this may be due to unknown contaminants, whereas in case of an ineffective treatment one cannot rule out the possibility that this is because the treatment was carried out inadequately [2,7,8]. Besides jeopardizing the *internal* validity and statistical power, insufficient treatment fidelity may also compromise the *external* validity of treatment outcome studies [2,5,7,9]. Notwithstanding its importance, several reviews demonstrated that most studies have not verified whether the fidelity of their treatments was adequate [2,3,7,10,11].

Various opinions exist on the number of components that treatment fidelity consists of. According to Perepletchikova and Kazdin [6], treatment fidelity consists of three components. These are *protocol adherence*, referring to the degree to which specific treatment procedures are used by the therapists during actual delivery of treatment [8], *competence*, which is the skillfulness of the therapists delivering the treatment [8], and *differentiation*, signifying whether the therapies differ from other treatments on several critical dimensions [12]. Others have added *treatment receipt* by

* Corresponding author. Tel.: +31 (0)45-5282753.
E-mail address: M.Leeuw@SRL.nl (M. Leeuw).

0895-4356/09/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi: 10.1016/j.jclinepi.2008.03.008

What is new?

- Failure to check treatment fidelity is a frequently observed threat to the internal and external validity of treatment outcome studies.
- Currently, there is a lack of systematic approaches for the verification of the fidelity of treatments under study.
- This paper provides a methodological illustration of how treatment fidelity can be assessed, using the *Method of Assessing Treatment Delivery (MATD).*
- MATD was applied in a randomized controlled trial of two cognitive behavioral treatments in chronic back pain patients, showing good treatment fidelity.
- We propose to use a similar approach in future treatment outcome studies.

the patient, for example, if the patient understands and is able to use the treatment skills, and a patient's *enactment* upon treatment, for example, whether the patient is able to actually implement the learned behavior in daily life [9–11]. Of all aspects, it may be most straightforward to assess actual delivery of treatment, because this can be assessed in a direct and objective manner. We therefore argue that treatment outcome studies at least verify whether the delivery of the treatments occurred as intended, although bearing in mind that this is only part of the full concept of treatment fidelity. The assessment of treatment delivery consists of verifying the occurrence of essential components (protocol adherence) and the nonoccurrence of prohibited protocol deviations (absence of treatment contamination) as well as verifying sufficient treatment differentiation [2,3,8,10].

Depending on the purpose, the manner in which treatment delivery can be assessed may differ. First, treatment delivery can be evaluated concurrently and repeatedly during the trial to improve this by providing feedback to the treatment agents. For this, it is necessary to reflect for *each individual treatment element* whether it is adequately carried out in case of required elements, and whether it is sufficiently absent in case of prohibited ones. By this means, treatment delivery can be optimized specifically with respect to the relevant elements [2,5–7,9]. Second, this assessment can be carried out after completion of the trial to determine whether treatment outcome comparisons are fair in the sense that both treatments were equally and adequately carried out according to their protocols. For this, it may be sufficient to appraise to *overall* treatment delivery for both treatments (e.g., the average protocol adherence, treatment contamination, and differentiation) in addition to verifying whether treatment delivery was equal between both treatments.

Even though some papers present general recommendations for the assessment of treatment delivery as well as other aspects of treatment fidelity [6–9], these remain rather indefinite. The studies indicating to have assessed treatment fidelity, did not describe in detail how this was evaluated, probably due to space limitations imposed by most journals [2,7]. Furthermore, the authentication of the reliability and validity of any assessment method is essential [3]. The aim of this study was to propose a Method of Assessing Treatment Delivery (MATD) based on crucial elements presented in the literature.[a] In this paper, we will apply this method to a completed trial in the field of behavioral medicine, to verify whether fair effectiveness comparisons were made between two multidisciplinary behavioral treatments. The feasibility, reliability, and content validity of MATD will be tested [3,6], and a comprehensive illustration of how the method was designed and applied will be given. Although the exact specifications of such a method are determined by the contents of the treatments of interest [6], the detailed description of our method might be helpful for other researchers in developing their own assessment tool. First, some relevant background information about the interventions is outlined.

2. Background information

There is accumulating evidence that in chronic low back pain, fear of pain is more disabling than pain itself [13,14]. It has therefore recently been chosen as an important target for intervention when aiming to reduce disability in these patients. Preliminary support was found for exposure in vivo (EXP) as an effective intervention in diminishing disability by reducing pain-related fear (e.g., [15–17]). In a multicenter randomized controlled trial (ISRCTN88087718), we compared the effectiveness of EXP with operant graded activity (GA) in chronic low back pain patients [18].

Both treatments aim at reducing functional disability, but EXP aims to achieve this by systematically reducing pain-related fear by gradually exposing patients to threatening and previously avoided activities [19,20], whereas GA aims to optimize active healthy behavior by encouraging patients to gradually increase their activity levels according to a time contingent treatment plan and by positively reinforcing healthy behavior [21].

Both treatments were carried out in four treatment centers by 19 therapist teams consisting of a psychologist and a physiotherapist or occupational therapist. All therapists had at least half a year of relevant clinical experience. The teams performed both treatments to ensure that the general

[a] We mainly relied on recommendations presented in existing literature. For these, we refer to the relevant source. When we encountered feasibility issues during designing and applying MATD for which we did not encounter any information in the literature, we added some personal undertakings to these. These will be recognized by the absence of a reference.

qualities of the therapists were evenly distributed across conditions. Both EXP and GA were highly structured and consisted of approximately 16 and 26 sessions, respectively. It is beyond the scope of this article to describe the efforts we undertook to enhance the fidelity of the treatments delivered, but several of these are also described by Moncher and Prinz [2] and Perepletchikova and Kazdin [6].

We found that EXP, despite its superior ability in reducing pain-related fear, was equally effective as GA in reducing functional disability and main complaints, although the difference between treatment conditions almost reached statistical significance favoring EXP [18].

3. Designing MATD

MATD needs to verify whether during delivery of treatment important treatment elements are actually addressed (maximal protocol adherence) while proscribed elements are not (minimal or no contamination), and whether the treatment can be differentiated from other treatments [2,3,6,8,10]. In case of comparing two interventions, some treatment components may apply to both treatment conditions, but other elements will be unique for a particular treatment and are therefore proscribed in the contrasting condition. The method has to represent a range of feasible elements defined by the treatment manual of each treatment, enabling for comparing the treatments on all of the discriminating as well as overlapping components [8].

We therefore developed a single MATD that aimed to determine protocol adherence and treatment contamination for EXP and GA simultaneously, together with differentiation between treatments. Evaluations of EXP sessions are expected to be reflected in positive scores on the unique and general required treatment elements of EXP, and in negative scores on those unique to GA (and thus proscribed for EXP), whereas the opposite applies for the GA sessions. The design of MATD consists of five steps, which are described below. In each step, first some general information is presented, after which its implementation is illustrated within our trial.

3.1. Step 1: Dividing the treatment into phases

3.1.1. General information

In the literature, we did not encounter any information or recommendations regarding the application of MATD with respect to different stages of treatment. Nevertheless, this is necessary when treatment content differs substantially between sessions, and thus when differential presence/absence of treatment elements is required for different phases of treatment. For example, the fact that a behavioral experiment is not yet performed during the explanation of the treatment rational does not imply that protocol adherence is insufficient, but rather that this element is only essential later on during treatment.

3.1.2. Illustration

In our study, based on the content of the sessions, we divided EXP and GA in three distinct phases: the *preparation phase*, that consisted of the establishment of a fear hierarchy in EXP, and of a patient's baseline activity level in GA; the *educational phase*, during which in both treatments the treatment rational was explained to the patient; and the *treatment phase*, that consisted of behavioral experiments in EXP, and of positive reinforcement of time contingent activity increases in GA.

3.2. Step 2: Identification of possible treatment elements

3.2.1. General information

MATD includes elements that could occur during treatment, comprising elements that are required, allowed, or not allowed during the treatments of interest [6,8,10]. In case of protocolized interventions, the identification of these elements should be rather straightforward. However, this can be complicated for trials with treatment as usual as the control condition, where guidelines are lacking to identify elements that are required or prohibited to occur. The prohibited treatment elements may not only consist of elements that are proscribed for the treatments of interest in general (e.g., giving a sole biomedical explanation of low back pain within a bio-psycho-social intervention), but also of those that are unique to the contrasting treatment (e.g., given that "performing a behavioral experiment" is clearly a unique EXP element, this is prohibited in GA). These elements have to be determined separately for each of the previously identified treatment phases. Obviously, these elements can best be identified by persons with profound knowledge of the treatments of interest. Detailed definitions of these elements may be useful to resolve any ambiguity about the meaning of a treatment component.

3.2.2. Illustration

In our study, two experts of both treatment protocols (J.W.V. and M.L.) jointly identified various possible treatment elements for EXP and GA, separately for the three treatment phases. The elements of GA and EXP were then intermixed and listed in a random order. This resulted in 16 items for the preparatory phase (Table 1), 24 items for the educational phase (Table 2), and 20 items for the treatment phase (Table 3). Unfortunately, we did not provide detailed definitions of these treatment elements at this stage.

3.3. Step 3: Categorization of the treatment elements

3.3.1. General information

As recommended, each of these previously identified treatment elements has to be classified into one of the following categories: (1) essential and unique; (2) essential but not unique; (3) compatible but not essential and not unique; (4) prohibited [6,8], to which we added a fifth category, that is (5) unique but not essential. Because it is important to

Table 1
The categorization and percentage of occurrences observed of the specific elements of the *preparatory phase* of MATD displayed separately for EXP and GA (five recorded sessions evaluated per treatment condition)

Specific elements of the preparatory phase	EXP category (% present)	GA category (% present)
Essential and unique EXP		
The patient's concern/fear with regard to activities is being discussed	EU (30)	P (10)
The patient assesses the level of perceived threat value of daily activities	EU (100)	P (0)
Photographs of daily activities are being used	EU (100)	P (0)
A hierarchy is being developed based on the threat value of daily activities	EU (100)	P (0)
Essential and unique GA		
The patient's baseline level of activities is being determined	P (0)	EU (50)
The patient performs activities in a pain-contingent manner	P (0)	EU (100)
The performance of activities is being recorded	P (0)	EU (90)
Important functional activities are being determined or performed	P (30)	EU (100)
Essential but not unique in EXP and GA		
There is good teamwork between the therapists and patient	E (100)	E (100)
The aim of the current session is being explained to the patient	E (70)	E (90)
The therapists respond understandingly to the problems expressed by the patient	E (100)	E (100)
Compatible in EXP and GA		
Inquiries are made about the patient's feelings and mood	C (10)	C (20)
Prohibited in EXP and GA		
The therapists go into possible medical causes of the symptoms	P (0)	P (0)
The therapists use medical terminology (e.g., diagnostic labels)	P (0)	P (0)
The therapists express fear/concern with regard to pain or activities	P (0)	P (0)
Other categorizations		
Supportive activities are being determined or carried out	P (0)	U (80)

Abbreviations: EXP, exposure in vivo; GA, graded activity; EU, essential and unique; E, essential but not unique; U, unique but not essential; C, compatible but not essential and not unique; P, prohibited.

Note: These specific treatment elements are presented in a *random* order in MATD.

evaluate the validity of such method [6], we suggest that two experts of the treatment protocols independently allocate these elements to these previously mentioned categories. Sufficient agreement between these experts in categorizing these elements can be conceived as support for the content validity of MATD.

3.3.2. Illustration

In our study, the two experts who earlier identified the feasible treatment elements, now independently assigned these to the five categories ranging from "essential and unique" to "prohibited," for the three phases for EXP and GA separately. Thus, different categorizations were collected for each phase of each treatment condition.

3.4. Step 4: Establishing the content validity of MATD

3.4.1. General information

Adequate agreement between experts in allocating the treatment elements to the same category can be verified by calculating Cohen's kappa. For the elements for which conformity is not found, agreement may either be achieved by consultation between these experts, or these may be removed from MATD.

3.4.2. Illustration

In our study, the agreement between the experts in categorizing the treatment elements was adequate (Cohen's kappa = 0.73). For example, the experts agreed that "a behavioral experiment is being performed" and "the level of activities is being increased in a time-contingent manner" are essential-and-unique elements for the treatment phase of EXP and GA, respectively. For the other items, they reached consensus after a single consultation. The final agreed categorizations of the treatment elements are displayed in Tables 1–3.[b]

3.5. Step 5: Making a scoring form for MATD

3.5.1. General information

A scoring form for MATD can be constructed, listing all of the previously determined treatment elements [6,8,10], behind which it can be indicated whether it did or did not occur during treatment [6,22]. It may also comprise a question as to which treatment condition the treatment session belongs [10].

3.5.2. Illustration

In our study, per phase all treatment elements were listed in a random order without revealing whether these elements were essential/essential and unique/unique but not

[b] It should be noted that, while unique elements should not be allowed in the contrasting treatment, the experts categorized two items to be unique to one treatment, but compatible to the other (one item of the educational and treatment phase).

Table 2
The categorization and percentage of occurrences observed of the specific elements of the *educational phase* of MATD displayed separately for EXP and GA (four recorded sessions evaluated per treatment condition)

Specific elements of the educational phase	EXP category (% present)	GA category (% present)
Essential and unique EXP		
The patient's concern/fear with regard to activities is being discussed	EU (63)	P (0)
It is explained that the treatment is aimed at verifying examining cognitions	EU (75)	P (0)
The circular model pain—pain cognitions—avoidance—pain is being explained	EU (88)	P (0)
Essential and unique GA		
The circular model of pain-inactivity-pain is being explained	C (13)	EU (75)
Examples of positive reinforcements are being identified	P (0)	EU (75)
A fluctuating pain-contingent activity pattern is being discussed	P (13)	EU (100)
It is explained that the treatment is aimed at ending inactivity	P (13)	EU (38)
Essential but not unique in EXP and GA		
It is emphasized that in chronic pain no clear relationship exists between pain and injury	E (0)	E (50)
There is good teamwork between the therapists and patient	E (100)	E (88)
The aim of the current session is being explained to the patient	E (88)	E (75)
A biomedical approach to pain is being discouraged	E (0)	E (50)
It is emphasized that pain reduction is not a therapy goal	E (88)	E (75)
The patient is being actively involved in the explanation of the therapy	E (100)	E (100)
A bio-psycho-social approach to pain is being explained	E (25)	E (0)
It is emphasized that all activities are possible	E (38)	E (13)
The drawbacks of inactivity are being explained	E (25)	E (63)
The therapists respond understandingly to the problems expressed by the patient	E (100)	E (100)
It is explained that the aim of the therapy is an increase in activity level	E (50)	E (75)
Compatible in EXP and GA		
There is an understanding attitude with regard to the patient's behavior	C (100)	C (100)
The patient's motivation for the therapy is being checked	C (100)	C (63)
Inquiries are made about the patient's feelings and mood	C (38)	C (13)
Prohibited in EXP and GA		
The therapists use medical terminology (e.g., diagnostic labels)	P (0)	P (0)
The therapists go into possible medical causes of the symptoms	P (0)	P (25)
The therapists express fear/concern with regard to pain or activities	P (0)	P (0)

Abbreviations: EXP, exposure in vivo; GA, graded activity; EU, essential and unique; E, essential but not unique; U, unique but not essential; C, compatible but not essential and not unique; P, prohibited.

Note: These specific treatment elements are presented in a *random* order in MATD.

essential/compatible/prohibited. After each treatment element, a dichotomous response choice was presented, so that its occurrence or nonoccurrence during treatment delivery could be indicated. MATD also included the question "which treatment do you believe that the current session belongs to?" together with response choices listing these treatment conditions (Fig. 1). We added detailed definitions of the specific treatment elements to MATD, so that the elements were unambiguous to future users completing MATD. As can be seen in Tables 1–3, there is considerable overlap between both treatment conditions on several of the treatment elements: in total, all essential-but-not-unique items were applicable to both treatments, in addition to 5 compatible and 10 prohibited elements. However, MATD also comprised treatment elements that were unique to EXP and thus proscribed for GA, and vice versa.

4. Applying MATD: Evaluating treatment delivery

The application of MATD in evaluating treatment delivery consists of five steps, which are presented and illustrated below.

4.1. Step 1: Recording of treatment sessions

4.1.1. General information

Adequate representations of treatment delivery are best collected concurrently with delivery of treatment. Audio recordings are suitable and preferred over self-report measures of patients or therapists, because these are objective reflections of actual treatment delivery [2,6,9]. Videotaping might even be more suitable, because this provides information about nonverbal communication that will not be available otherwise. Treatment sessions ought to be recorded as many as possible, from which a random selection can be drawn for actual assessment [6]. The random selection ideally reflects the variation in sessions and patients, to optimize generalization of the findings throughout treatment and among patients [2,6,7].

4.1.2. Illustration

We provided each treatment center with an MP3 recording device. Out of approximately 1,275 treatment sessions performed, 265 (21%) were recorded. The median number of recorded sessions per therapist team was 5 (SD = 14.96,

Table 3
The categorization and percentage of occurrences observed of the specific elements of the *treatment phase* of MATD displayed separately for EXP and GA (six recorded sessions evaluated per treatment condition)

Specific elements of the treatment phase	EXP category (% present)	GA category (% present)
Essential and unique EXP		
The patient's concern/fear with regard to activities is being discussed	EU (92)	P (0)
A catastrophizing cognition is being identified	EU (83)	P (0)
A behavioral experiment is being performed	EU (75)	P (0)
Activities from the hierarchy or based on threat value are being performed	EU (67)	P (0)
Clear agreements are made about the way in which an activity should be carried out (e.g., how often, how high the jumps should be, how to bend down)	EU (42)	C (17)
Essential and unique GA		
The level of activities is being increased in a time-contingent manner	P (8)	EU (75)
Activities are being carried out according to a time-contingent plan of treatment	P (8)	EU (75)
There is positive reinforcement of activity quotas that are met	P (0)	EU (8)
Essential but not unique in EXP and GA		
Homework is being assigned	E (58)	E (25)
The aim of the current session is being explained to the patient	E (0)	E (33)
There is good teamwork between the therapists and patient	E (92)	E (100)
The therapists respond understandingly to the problems expressed by the patient	E (100)	E (100)
Homework is being evaluated	E (67)	E (58)
Compatible in EXP and GA		
Inquiries are made about the patient's feelings and mood	C (33)	C (50)
Prohibited in EXP and GA		
The therapists use medical terminology (e.g., diagnostic labels)	P (0)	P (0)
The therapists express fear/concern with regard to pain or activities	P (0)	P (8)
The therapists pay a lot of attention to the patient's pain behavior	P (50)	P (8)
The therapists go into possible medical causes of the symptoms	P (0)	P (0)
Other categorizations		
A catastrophizing cognition is being evaluated	U (67)	P (0)
The performance of activities is being recorded	C (8)	E (67)

Abbreviations: EXP, exposure in vivo; GA, graded activity; EU, essential and unique; E, essential but not unique; U, unique but not essential; C, compatible but not essential and not unique; P, prohibited.

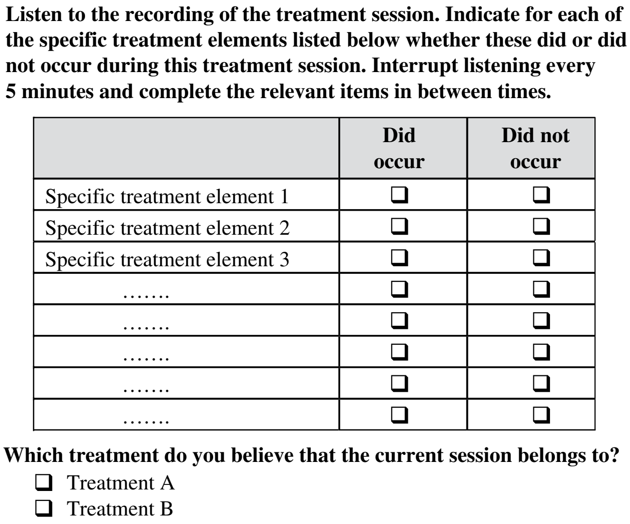
Note: These specific treatment elements are presented in a *random* order in MATD.

min = 0, max = 47). Subsequently, we drew a random selection of 30 sessions (11% of the available recordings) from these to be rated for actual treatment delivery, because statistically this number of sessions was minimally required to calculate interrater reliability. It was taken care of that for each team of therapists both an EXP and a GA session were selected if possible, that these sessions were derived from all three treatment phases, that a session was completely recorded, and that the quality of recording was sufficient to understand the conversation. Unintelligible recordings were replaced by another randomly selected one until these criteria were met. Eventually, this resulted in, respectively, five, four, and six recordings for the preparatory, educational, and treatment phase of EXP and GA.

4.2. Step 2: Selecting raters

4.2.1. General information

When intended to verify the interrater reliability of MATD, at least two raters are needed, who independently rate the selected treatments recordings with this measure. It is important to keep these raters blind with regard to the study hypotheses and as independent of the study as possible [6,8,10]. These raters have to be presented with sufficient information about the treatment condition(s), and it may be helpful to accustom them to using MATD by practice [2,6]. In case of straightforward ratings, undergraduate students with an education relevant for the treatments of interest may be capable of assessing the



Listen to the recording of the treatment session. Indicate for each of the specific treatment elements listed below whether these did or did not occur during this treatment session. Interrupt listening every 5 minutes and complete the relevant items in between times.

	Did occur	**Did not occur**
Specific treatment element 1	❑	❑
Specific treatment element 2	❑	❑
Specific treatment element 3	❑	❑
…….	❑	❑
…….	❑	❑
…….	❑	❑
…….	❑	❑
…….	❑	❑

Which treatment do you believe that the current session belongs to?
❑ Treatment A
❑ Treatment B

Fig. 1. Example of a scoring form of MATD.

(non-)occurrence of treatment elements during treatment delivery. However, when the evaluation requires a certain expertise, experienced therapists are more suitable [8].

4.2.2. Illustration

In our study, two undergraduate students in clinical psychology were trained as raters, because we believed that the assessment of the presence or absence of the treatment elements was fairly straightforward. These students had no previous involvement in the study and were kept blind with regard to the study hypotheses, except for their involvement in the assessment of interrater reliability. They were paid for the training and for the actual scoring of the sessions. They first gained knowledge of both treatments by reading the detailed treatment manuals. Then, they practiced three sessions together with one of the developers of MATD (M.L.). After this, each student rated the 30 selected recordings of treatment sessions.

4.3. Step 3: Scoring of the treatment sessions

4.3.1. General information

We did not encounter any recommendations in the literature with regard to the scoring of the treatment sessions.

4.3.2. Illustration

To ensure adequate ratings, we gave our raters the instruction to interrupt their listening to the recordings every 5 minutes, and to fill in MATD for the relevant elements in between times. Consultation between raters was prevented.

4.4. Step 4: Calculation of interrater reliability

4.4.1. General information

The interrater reliability between the raters [2,3] can be determined by calculating Cohen's kappa, both for the specific treatment elements and the treatment categorization. Providing adequate interrater reliability, the mean of the two raters can be taken to calculate the mean proportion of essential or prohibited treatment elements.

4.4.2. Illustration

In our study, the agreement found between both raters, as represented by Cohen's kappa, was 0.72 for the specific treatment elements and 0.87 for the categorization of treatment condition, indicating good interrater reliability of MATD.

4.5. Step 5: Verifying treatment delivery

4.5.1. Assessing protocol adherence

General information: A priori, it has to be determined to what degree required treatment elements should have occurred during treatment, to consider protocol adherence as adequate [6]. As far as we know, there are no clear guidelines available for this. Whereas Perepletchikova and Kazdin [6] propose that "high integrity levels may be represented by 80–100% integrity, whereas low integrity condition may be represented by 50% integrity or less" (p. 377), it remains unclear how the level of integrity was assessed. Moreover, they suggest that in multifaceted and time-consuming treatments, adherence ratings are expected to be lower than in uncomplicated interventions. Protocol adherence can be computed by dividing the number of observed required treatment elements (essential and unique and essential but not unique) by the maximum possible number of these elements [7]. Higher scores therefore indicate higher frequencies of essential treatment elements, and thus higher protocol adherence. Optionally, one can include the unique but not essential and compatible items into the assessment of protocol adherence, but because these are not required, it is difficult to determine to what degree these need to be addressed during treatment.

Illustration: In our study, we judiciously determined the cut-off point for sufficient protocol adherence, which we defined as the occurrence of at least 70% of essential treatment elements. First, we took into account the fact that both treatments were complex, because these comprised multiple intervention techniques, and were carried out by multiple multidisciplinary therapist teams in several treatment centers. Furthermore, we considered that under some circumstances essential treatment components could be less important (e.g., when behavioral experiments are omitted when catastrophic cognitions cannot be identified), or even irrelevant (e.g., when a session is dedicated to an important event the patient goes through).

The occurrences observed of the essential specific treatment elements are displayed in Tables 1–3. The mean protocol adherence scores, reflected in the mean proportion of essential treatment elements over all evaluated treatment sessions for the phases of EXP and GA, are displayed in Table 4. ANOVA, with protocol adherence as the dependent variable and treatment condition and phase as fixed factors (alpha = 0.05), demonstrated that the protocol adherence did not differ significantly between the treatment conditions for each of the phases ($F = 1.30$, $P = 0.28$), indicating that further analyses could be performed with the mean scores independent of phase. Furthermore, no difference was found in average protocol adherence between both treatment conditions ($F = 0.05$, $P = 0.82$). It was found that in general 72% (SD = 18.71) of the essential elements occurred during the selected treatment sessions, indicating that the preset criterion of good protocol adherence was met. However, it should be noted that protocol adherence was especially high during the preparatory phase (86–91%), whereas it dropped to 58–68% during the subsequent phases.

4.5.2. Assessing treatment contamination

General information: An a priori level of the degree in which prohibited treatment elements are allowed during treatment is needed to consider treatment contamination absent [6]. Again, to our knowledge, no guidelines are available for this. Treatment contamination can be computed by dividing the number of observed prohibited treatment elements by the maximum possible number of these elements, with higher scores indicating higher frequencies of prohibited treatment elements, and thus higher degrees of contamination.

Illustration: In our study, sufficiently low treatment contamination was defined as the occurrence of maximally 10% of the prohibited treatment elements during treatment. Because contamination sum scores per session were not normally distributed, with hardly any values larger than 0, both parametric and nonparametric testing were inappropriate. We therefore dichotomized the contamination sum scores by recoding scores above 0 into 1, and subsequently performed separate chi-square tests between treatment conditions in general or per phase of treatment (alpha = 0.01 because of multiple testing). The occurrences observed of the prohibited specific treatment elements are displayed in Tables 1–3. The mean contamination scores for the phases of EXP and GA are displayed in Table 4.

Contamination scores did not differ significantly between the treatment conditions for each of the phases (chi-square's <4.44, P-values >0.04), thus further analyses could be performed independent of phase of treatment. Also, contamination scores did not differ between both treatment conditions (chi-square = 3.07, $P = 0.08$). It was demonstrated that overall 4% (SD = 7.71) of the prohibited elements occurred during the evaluated treatment sessions, indicating that treatment contamination was sufficiently absent. For each of the treatment phases, the mean contamination score was below the predefined 10%.

4.5.3. *Assessing treatment differentiation*

General information: One way to verify treatment differentiation is by calculating the percentage of adequate treatment categorizations by the independent raters.

Illustration: In the absence of guidelines, we judiciously determined that more than 90% of the sessions had to be classified correctly. Both independent raters classified the recorded sessions to the correct treatment condition in 97% of the cases: each rater incorrectly classified one EXP session as GA. There was no significant difference in accurate categorizations between treatment conditions ($F = 2.07$, $P = 0.15$). Thus, both treatments could be easily differentiated from each other.

5. Conclusion and discussion

Based on available recommendations in the literature, we succeeded in designing, as well as feasibly applying, a method to assess treatment delivery in a randomized controlled trial comparing two behavioral treatments for chronic low back pain. Sufficient agreement between two experts in evaluating the specific items as being essential/essential and unique/unique but not essential/compatible/prohibited supported the content validity of MATD. The reliability of the application of MATD was supported by adequate interrater reliability of two independent raters. From their ratings of a selection of recorded treatment sessions, it was shown that protocol adherence was sufficient, that treatment contamination was almost absent, and that treatment differentiation was substantial, at least according to our a priori criteria. Besides the importance of establishing adequate treatment delivery in both conditions, another relevant finding was that treatment delivery was equal between both treatments. This result at least ensures that neither of these treatment conditions was favored by being better conducted as compared to the other. The comparisons between the effectiveness of EXP as compared to GA therefore seem fair.

Whereas this study put a great effort in quantitatively verifying treatment delivery, this procedure should not be interpreted without acknowledging some drawbacks. There are four main limitations that need mentioning. First, it is difficult to predefine criteria for adequate treatment delivery in the absence of clear guidelines. We found it challenging to predetermine how many essential elements should have occurred for adequate protocol adherence, especially considering that under some circumstances essential treatment components could be less important or even irrelevant (see Section 4.5.1). Furthermore, retrospectively, we may even have overestimated the importance of several items. For

Table 4

Observed protocol adherence (mean proportion of essential treatment elements), observed treatment contamination (mean proportion of prohibited treatment elements), and observed treatment differentiation (percentage of correct treatment classifications) as a reflection of treatment delivery for the preparatory, educational, and treatment phases of EXP and GA

	Protocol adherence: mean proportion (%) of essential treatment elements (SD)		Treatment contamination: mean proportion (%) of prohibited treatment elements SD)		Treatment differentiation: correct treatment classification (%)	
	EXP	GA	EXP	GA	EXP	GA
Preparatory phase	86.25 (9.83)	90.83 (11.08)	3.00 (4.83)	1.25 (3.95)	100.00	100.00
Educational phase	65.34 (11.18)	67.19 (19.55)	4.17 (11.79)	4.17 (7.72)	87.50	100.00
Treatment phase	67.50 (19.13)	58.33 (15.08)	9.03 (9.87)	2.08 (4.87)	91.70	100.00
Total	73.17 (16.97)	71.53 (20.55)	5.72 (9.24)	2.36 (5.44)	93.30	100.00

Note that no difference was found between treatment conditions in protocol adherence ($F = 0.05$, $P = 0.82$), treatment contamination (chi-square = 3.07, $P = 0.08$), or treatment differentiation ($F = 2.07$, $P = 0.15$).

example, during the education phase, the low occurrence of items 1, 9, 14, and 17 may be explained by the fact that these have become redundant because the rehabilitation physician previously explained the bio-psycho-social view of chronic pain, which was in fact in accordance with both our treatment protocols. For future designs of MATDs, we recommend to select only those specific ingredients that are essential irrespective of circumstances. Furthermore, we did not integrate the compatible or unique-but-not-essential treatment elements in the assessment of treatment delivery, because it is difficult to determine to what degree these need to be addressed during treatment. Forthcoming studies may individually decide whether these treatment elements are relevant for their assessment of treatment delivery.

Second, depending on the aim of the evaluation of treatment delivery, the manner in which its components are assessed and evaluated differs. Because it was our aim to evaluate whether treatment comparisons were fair, in the sense that both treatments were adequately and equally delivered, we averaged protocol adherence and treatment contamination over all evaluated sessions. However, by doing so, we disregarded the variability in treatment delivery between sessions and between treatment elements. For example, although in general protocol adherence was regarded as adequate, it was especially high during the preparatory phase, whereas it dropped below the predefined criterion during the subsequent phases. Furthermore, although the therapists adhered to some treatment elements in almost all of the cases, this was insufficient in case of other treatment elements. Of course, the currently presented MATD is also suitable for administration and evaluation concurrently (online) with a treatment outcome study. When the aim is to improve and optimize treatment delivery, these aspects should not only be assessed repeatedly *during* the trial, but also detailed information about adherence and contamination in each session and per treatment element will be essential.

Third, the mere assessment of treatment delivery disregards the evaluation of treatment competence [8], treatment receipt by the patient, and enactment of the patient upon treatment [9,11]. To fully understand, and compare, the effects or treatments, these components have to be encouraged and evaluated as well. Nevertheless, completion of the Treatment Fidelity Checklist [10] revealed that our study applied 17 out of 25 (68%) treatment fidelity strategies, which is higher than the reported average 55% of other studies: we used 100% of the treatment fidelity strategies of treatment design, 75% of training providers, and 80% of delivery of treatment, but none of those from receipt of treatment and enactment of treatment skills.

Fourth, therapist's awareness of the recording of their treatment sessions may influence their behavior. It is therefore possible that treatment delivery is enhanced in case of the sessions observed, and that because of this ratings are inflated [2,3,6,7]. In fact, in our study this might even be more of an issue, because only 21% of the treatment sessions were recorded despite repeated instructions and reminders to record as many treatment sessions as possible.

Despite these limitations, this study presents several important steps derived from available recommendations in the literature that can be undertaken when developing and applying MATD, and comprehensively illustrates these subsequent steps within a randomized controlled trial in the area of behavioral medicine. Presentations of methodological issues and detailed examples of these methods may encourage the verification of treatment delivery as a partial reflection of treatment fidelity in treatment outcome studies. Researchers preparing forthcoming studies may use these to develop their own MATD, although these methods obviously have to be modified according to the contents of the treatments of interest, and the aims of the assessment.

Acknowledgments

This study was supported by Grant nr. 1436.0002 of the Netherlands Organization for Health Research and Development (ZonMw). We are grateful to the many therapists and support staff from the treatment centers of Maastricht, Hoensbroek, Sittard, and Eindhoven, who performed the treatments and recorded the treatment sessions, to the patients willing to participate, to Sita van Riet for her assistance in the data collection, and to Sarah Holla and Dominique Gielen for their evaluations of treatment delivery.

References

[1] Yeaton WH, Sechrest L. Critical dimensions in the choice and maintenance of successful treatments: strength, integrity, and effectiveness. J Consult Clin Psychol 1981;49(2):156–67.

[2] Moncher FJ, Prinz RJ. Treatment fidelity in outcome studies. Clin Psychol Rev 1991;11:247–66.

[3] Peterson L, Homer AL, Wonderlich SA. The integrity of independent variables in behavior analysis. J Appl Behav Anal 1982;15:477–92.

[4] Mayo-Wilson E. Reporting implementation in randomized trials: proposed additions to the consolidated standards of reporting trials statement. Am J Public Health 2007;97:630–3.

[5] Cook TD, Campbell DT. Quasi-experimentation: design & analysis issues for field settings. Chicago, IL: Rand McNally College Publishing Company; 1979.

[6] Perepletchikova F, Kazdin AE. Treatment integrity and therapeutic change: issues and research recommendations. Clin Psychol Sci Pract 2005;12(4):365–83.

[7] Gresham FM, Gansle KA, Noell GH. Treatment integrity in applied behavior analysis with children. J Appl Behav Anal 1993;26:257–63.

[8] Waltz J, Addis ME, Koerner K, Jacobson NS. Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. J Consult Clin Psychol 1993;61:620–30.

[9] Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, et al. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH behavior change consortium. Health Psychol 2004;23(5):443–51.

[10] Borrelli B, Sepinwall D, Ernst D, Bellg AJ, Czajkowski S, Breger R, et al. A new tool to assess treatment fidelity and evaluation of

treatment fidelity across 10 years of health behavior research. J Consult Clin Psychol 2005;73:852−60.

[11] Lichstein KL, Riedel BW, Grieve R. Fair tests of clinical trials: a treatment implementation model. Adv Behav Res Ther 1994;16: 1−29.

[12] Kazdin AE. Comparative outcome studies of psychotherapy: methodological issues and strategies. J Consult Clin Psychol 1986;54: 95−105.

[13] Vlaeyen JWS, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. Pain 2000;85: 317−32.

[14] Leeuw M, Goossens MEJB, Linton SJ, Crombez G, Boersma K, Vlaeyen JWS. The fear avoidance model of musculoskeletal pain: current state of scientific evidence. J Behav Med 2007;30(1):77−94.

[15] Boersma K, Linton SJ, Overmeer T, Jansson M, Vlaeyen JWS, de Jong J. Lowering fear-avoidance and enhancing function through exposure in vivo; a multiple baseline study across six patients with back pain. Pain 2004;108(1−2):8−16.

[16] Vlaeyen JWS, de Jong J, Geilen M, Heuts PHTG, van Breukelen G. The treatment of fear of movement/(re)injury in chronic low back pain: further evidence on the effectiveness of exposure in vivo. Clin J pain 2002;18(4):251−61.

[17] de Jong JR, Vlaeyen JWS, Onghena P, Goossens MEJB, Geilen M, Mulder H. Fear of movement/(re)injury in chronic low back pain: education of exposure in vivo as mediator to fear reduction? Clin J Pain 2005;21(1):9−17.

[18] Leeuw M, Goossens MEJB, van Breukelen GJP, de Jong JR, Heuts PHTG, Smeets REJM, et al. Exposure in vivo versus operant graded activity in chronic low back pain patients: results of a randomized controlled trial. Pain, in press. doi:10.1016/j.pain.2007.12.009.

[19] Vlaeyen JWS, de Jong J, Leeuw M, Crombez G. Fear reduction in chronic pain: graded exposure in vivo with behavioral experiments. In: Asmundson GJ, Vlaeyen JWS, Crombez G, editors. Understanding and treating fear of pain. New York: Oxford University Press; 2004.

[20] Vlaeyen JWS, de Jong J, Sieben JM, Crombez G. Graded exposure in vivo for pain-related fear. In: Turk DC, Gatchel RJ, editors. Psychological approaches to pain management. A practitioner's handbook. New York: The Guilford Press; 2002.

[21] Sanders SH. Operant conditioning with chronic pain: back to basics. In: Turk DC, Gatchel RJ, editors. Psychological approaches to pain management. New York: The Guilford Press; 2002.

[22] Gresham FM. Treatment integrity and therapeutic change: commentary on Perepletchikova and Kazdin. Clin Psychol Sci Pract 2005;12(4):391−4.