ASSESSMENT OF RESEARCH CRITERIA FOR EXPOSURE-BASED OUTCOME STUDIES OF PTSD

JACQUES WILLIAM TAYLOR

Assignment presented in partial fulfilment of the requirements for the degree of Master of Arts (Counselling Psychology) at the University of Stellenbosch

Supervisor: Prof. A.T. Möller

December 2004

DECLARATION

I, the undersigned, hereby declare that the work contained in this assignment is my own original work, and that I have not previously in its entirety or in part submitted it at any university for a degree.

J.W. Taylor Date

SUMMARY

Treatment outcome research strives towards objective estimates of disorder-specific treatment efficacy and has been applied to most psychiatric disorders. However, due to shortcomings in outcome research designs, problems still remain regarding the interpretation and generalisability of treatment outcomes. This is despite the development of research methodology criteria such as the Gold Standards, currently viewed as essential criteria for well-controlled cognitive-behavioural outcome research. The objectives of this assignment are (a) to assess the Gold Standards as criteria for treatment outcome research by means of a qualitative overview and evaluation of exposure treatment studies for PTSD, and (b) to make recommendations for the expansion and/or modification of these criteria. An assessment of five selected treatment outcome trials, based on the Gold Standards, showed significant limitations in the scope of the Gold Standards regarding (a) the inclusion of target symptoms in the research hypotheses, (b) estimates of treatment adherence, (c) guidelines for statistical analyses of attrition points, (d) the ethical implementation of exposure treatment, and (e) estimates of significant clinical change. It is concluded that the Gold Standards are not sufficient to ensure valid and reliable treatment outcomes. Recommendations are made for the expansion of four of the existing Gold Standards parameters and three additional criteria are proposed.

Key words: treatment outcome research, Gold Standards, post-traumatic stress disorder, exposure therapy.

OPSOMMING

Navorsing oor behandelingsuitkoms streef na objektiewe resultate oor die uitkoms van behandeling vir spesifieke psigiatriese versteurings. Nogtans, weens tekortkominge in die ontwerp van uitkomsstudies, word talle probleme steeds ervaar met die interpretasie en veralgemeenbaarheid van die resultate van die studies. Dit is die geval ten spyte van die ontwikkeling van navorsingskriteria soos die "Gold Standards" wat huidig as die belangrikste kriteria vir uitkomsstudies op die gebied van die kognitiewe gedragsterapie aanvaar word. Hierdie projek het ten doel om (a) die Gold Standards as kriteria vir uitkomsnavorsing te assesseer deur middel van 'n kwalitatiewe oorsig en evaluering van vyf geselekteerde uitkomsstudies van blootstellingsterapie vir post-traumatiese stresversteuring, en (b) om aanbevelings te maak ter aanvulling en/of wysiging van die Gold Standards. Evaluasie van die studies het betekenisvolle beperkings in die Gold Standards se omvattenheid uitgelig in terme van (a) die insluiting van teikensimptome in die navorsingshipoteses, (b) die skatting van behandelingvoltrekking ("treatment adherence"), (c) riglyne vir die statistiese analise van data oor attrisie, (d) die etiese implementering van blootstellingsterapie, en (e) skattings van betekenisvolle kliniese verandering. Dit blyk dat die Gold Standards nie voldoende is om geldige en betroubare resultate oor behandelingsuitkomste te verseker nie. Aanbevelings word gemaak vir die hersiening van vier van die Gold Standards kriteria en drie addisionele kriteria word voorgestel.

Sleutelwoorde: behandelingsuitkomsnavorsing, Gold Standards, post-traumatiese stresversteuring, blootstellingsterapie.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to:

- Professor André Möller, for his invaluable support and interest in my personal development over the past years, as well as his professional contributions to the presentation of this assignment.
- My Parents, for giving me the opportunity to receive a tertiary education, and for their continuous support and belief in my abilities.
- Jacqueline, my sister, for her love, support, inspiration, and undying belief in me.
- All my friends, from whom I have received all the motivation one can ever ask for.

LIST OF TABLES

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 Outcome-based Research

In the United States, managed care insurance companies are increasingly insisting on therapeutic services that are based on procedures with proven efficacy (Hickling & Blanchard, 1997). The same is happening in South Africa, manifested through medical aid schemes that are cutting back on benefits for psychiatric disorders (Brümmer, 29 October 2003). The importance of providing empirical proof of the efficacy of psychological treatments is evident.

Apart from a pragmatic approach to determine treatment efficacy (Nathan & Gorman; and Williams & Sommer, cited in Wilson, Friedman, & Lindy, 2001), the empirical approach of finding research and practice support for psychological treatment options is enjoying increasingly wider support under researches (Wilson et al., 2001). Outcome-based research is an example of such empirical approaches (Agency for Healthcare Research and Quality [AHRQ], 2000).

Outcome-based research seeks to understand the results of particular health care practices and interventions, including psychotherapeutic interventions. End results include effects that people experience, such as a change in psychological symptoms and the ability to function more effectively. By linking the type of interventions patients receive to the outcomes they experience, outcome research has become the key to developing better ways to monitor and improve the quality of interventions. It provides clinicians and patients with evidence about benefits, risks, and results of treatments so they can make more informed decisions about their healthcare (AHRQ, 2000).

Treatment outcome research, particularly from a cognitive-behavioural perspective, has been applied to various psychiatric disorders (Caballo, 1998). Examples include major depression (Salkovskis, Atha, & Storer, 1990), agoraphobia (Salkovskis, Clark, & Hackmann, 1991), general anxiety disorder (Borkovec & Costello, 1993), social phobia (Turner, Beidel, & Cooley-Quille, 1995), obsessive-compulsive disorder (Emmelkamp, Visser, & Hoekstra, 1988), and post-traumatic stress disorder (PTSD) (Foa et al., 1999).

1.2 Gold Standards

The development of appropriate methodology to investigate treatment outcome efficacy has become a challenge. The Gold Standards (Foa & Meadows, 1997) have emerged as the most widely used set of criteria for methodologically rigorous treatment outcome research (Foa, Keane, & Friedman, 2000; Foa & Meadows, 1997; Wilson et al., 2001). The Gold Standards were developed, under the auspices of the International Society for Traumatic Stress Studies (ISTSS) (Foa et al., 2000), as part of a larger project to formulate treatment guidelines for PTSD. The Gold Standards comprise of seven parameters for planning the "ideal" outcome study (Foa & Meadows, 1997, p. 455).

a) Clearly defined target symptoms

The target syndrome (e.g., PTSD) or target symptom (e.g., distressing intrusive recollections) must be clearly defined, to aid in diagnosis so that appropriate intervention measures are selected. It is also important to specify a threshold of symptom severity as an inclusion criterion for participants to minimize bias in the interpretation of results. Finally, related to target symptoms, specific inclusion and exclusion criteria must be delineated. This assists in the examination of outcome predictors and the relative efficacy of treatment regardless of sample differences.

b) Reliable and valid measures

Once the population and target symptoms have been identified, assessment measures with reliable and valid psychometric properties must be included. Studies that investigate specific diagnoses must include measures designed to yield diagnoses as well as instruments that assess symptom severity.

c) Blind evaluators

In order to minimize expectancy and demand bias, the use of blind evaluators is crucial, and entails that the assessor of treatment outcome not be the person conducting treatment and, that participants be trained not to reveal their treatment condition at post-treatment and follow-up assessment.

d) Assessor training

The reliability and validity of assessment of treatment outcome depends largely on the skill of the evaluator, and therefore necessitates the training of assessors. The minimum criteria required include demonstrating interrater reliability and calibrating assessment procedures during the study, in order to avoid evaluator drift.

e) Specific, manualised and replicable treatments

Specific treatments must be employed, designed to address the target syndrome. To ensure consistent treatment delivery, the specific treatments must be detailed in treatment manuals, which will afford replicability and contribute towards outcome generalisability.

f) Unbiased assignment of participants to treatments

To help ensure that observed treatment differences or similarities are due to the techniques employed and not to extraneous factors, participants should be randomly assigned to treatment conditions or by way of a stratified sampling approach. To separate the effects of treatment from therapists, each treatment should be delivered by at least two therapists, and participants should be randomly assigned to therapists within each condition.

g) Treatment adherence

Treatment adherence ratings aim to inform whether or not treatments were carried out according to protocol and whether certain components of one treatment drifted into another. These ratings are of importance when one treatment excludes a technique that is part of competing treatment and the aim of the study is to evaluate the importance of that technique.

The Gold Standards have not formally been evaluated in an empirical and/or review study. Harvey, Bryant and Tarrier (2003) and Foa et al. (2000) only stated the content and value of the Gold Standards. Maxfield and Hyer (2002) used the Gold Standards in an eye movement desensitisation and reprocessing (EMDR) study to investigate the relationship between methodology and treatment efficacy, and proposed modifications to the Gold Standards to suit EMDR research. According to Foa and Meadows (1997), the Gold Standards are "...by no means exhaustive" (p. 474), and conceded that guidelines for statistical procedures must be included.

Despite the importance of conducting treatment outcome research and the development of research guidelines, such as the Gold Standards, treatment outcome ambiguities regarding the efficacy of various psychological treatments still remain (Wilson et al., 2001). To illustrate this research dilemma, PTSD and its cognitive-behavioural treatment are discussed in the next section.

1.3 Post-Traumatic Stress Disorder

1.3.1 Background

Post-traumatic stress disorder is a debilitating psychiatric disorder that can affect up to 30% of people exposed to a life threatening stressor (Foa & Meadows, 1997), with a general population prevalence of up to 12% (Sherman, 1998). The Diagnostic and Statistical Manual for Mental Disorders (DSM-IV-TR) (American Psychiatric Association [APA], 2002) includes six criteria in the diagnosis of PTSD. *Firstly,* there must be exposure to an extreme traumatic stressor involving direct personal experience, witnessing or learning about an event that involves actual, threatened death, or serious injury. *Secondly*, the person's response to this event must involve intense fear, helplessness, or horror. The characteristic symptoms resulting from the exposure to the trauma include, in the *third* place, persistent re-experiencing of the traumatic event; *fourthly*, persistent avoidance of stimuli associated with the trauma and numbing of general responsiveness, and *fifthly,* persistent symptoms of increased arousal. *Finally*, the full symptom picture must be present for more than one month, and the disturbance must cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Since its incorporation as a diagnostic category in the DSM-III (APA, 1980), there has been a considerable interest in PTSD as a diagnostic category (Foa & Meadows, 1997). This is evident from more than 22 000 publications on PTSD up and till September 2003, according to the "PsychInfo" database.

1.3.2 Cognitive-behavioural treatment of PTSD

PTSD as a new diagnostic category provided the impetus for the development of cognitive-behavioural treatment programs (Wilson et al., 2001). Various cognitive-behavioural approaches have been initiated, with varying degrees of effectiveness. Most notably are the exposure-based interventions, including systematic desensitisation (Frank et al., 1988), imaginal and *in vivo* exposure treatment (Foa & Kozak, 1986), and EMDR (Carlson, Chemtob, Rusnak, Hedlund, & Muraoka, 1998; Shapiro, 1989), and the anxiety management programs, including stress inoculation training (SIT) (Foa, Rothbaum, Riggs, & Murdock, 1991; Foa, Rothbaum, & Steketee, 1993), and cognitive processing therapy (CPT) (Resick & Schnicke, 1992).

*1.3.3 Imaginal and **in vivo** exposure therapy*

In the treatment of PTSD, any of the exposure-based treatment approaches can be used to access emo-

tions associated with the traumatic event and to promote emotional processing. Experts view emotional processing as the essential ingredient for treating PTSD (Foa & Kozak, 1986). It is however imaginal and *in vivo* exposure treatment that enjoys widespread support in the literature (Harvey et al., 2003).

Imaginal and *in vivo* exposure treatment typically involve some form of graduated exposure to trauma-relevant cues, a procedure that may or may not be accompanied by attempts to maintain a fear-antagonistic state such as relaxation. Exposure could include returning to the place where the traumatic event occurred and attempting to come into contact with some of the salient contextual cues (e.g., time of day that the traumatic incident occurred), as well as contact with other stimuli that have affective associations to the event (e.g., sounds or odours). More often, imaginal material is presented alone or in combination with *in vivo* cues in order to help the individual access the full array of emotional structures associated with a particular traumatic event (Foa & Meadows, 1997).

"Pure[1]" exposure therapy is currently viewed as a "well-established treatment" (Rothbaum & Schwartz, 2002, p. 59) and classified by the AHRQ as a level A treatment option for PTSD (Foa et al., 2000), with support from controlled treatment outcome studies (Boudewyns, Hyer, Woods, Harrison, & McCranie, 1990; Bryant, Moulds, Guthrie, Dang, & Dixon, 2003; Cooper & Clum, 1989; Fairbank & Keane, 1982; Foa et al., 1991, 1999; Keane, Fairbank, Caddell, & Zimering, 1989; Keane & Kaloupek, 1982; Marks, Lovell, Noshirvani, Livanou, & Thrasher, 1998; Paunovic & Öst, 2001; Resick, Nishith, Weaver, Astin, & Feuer, 2002; Tarrier et al., 1999; Taylor et al., 2003).

Although exposure therapy is considered as the psychological treatment of choice for PTSD (Foa et al., 2000; Foa & Meadows, 1997; Rothbaum & Schwartz, 2002), empirical results indicate that cognitive-behavioural treatment approaches, such as cognitive restructuring, play an equal or even superior role in the amelioration of PTSD symptoms (Resick et al., 2002). These results raises questions regarding the validity of exposure therapy's status, and question the methodology employed to reach these treatment outcome results.

1.4 Motivation, Problem Statement and Goals

Treatment outcome research, and in particular its methodology, have become important in determining

---

[1] "Pure" exposure treatment may contain imaginal and/or *in vivo* exposure, the use of a "subjective units of distress " (SUD) scale, and secondary relaxation training (e.g. breathing retraining) as an adjunct to the exposure components.

optimal treatment options. The Gold Standards are accepted as essential criteria for treatment outcome research. However, as indicated in section 1.3, results on the outcome of different treatment modalities warrants an investigation of the methodologies employed in these outcome studies, particularly of the Gold Standards as essential research criteria in treatment outcome research.

The objectives of this assignment are therefore (a) to assess the Gold Standards as criteria for treatment outcome research by means of an overview and evaluation of selected exposure studies for PTSD, and (b) to make recommendations for the expansion and/or modification of these criteria, if appropriate.

CHAPTER 2
METHODOLOGY

2.1 Methodology

As stated in section 1.4, this assignment investigates the scope of the Gold Standards in order to determine whether it represents sufficient criteria for valid and reliable treatment outcome research.

Consequently, an overview of selected exposure outcome studies for PTSD, adhering to the Gold Standards, is provided. The results of these studies are then assessed in terms of the methodology employed, in order to identify possible limitations and shortcomings regarding methodology, particularly as it pertains to the Gold Standards. Finally, recommendations are made to expand and/or modify the Gold Standards.

2.2 Literature Search

The PsychInfo and South African Studies databases were searched for exposure treatment outcome studies on PTSD, using the following key-words and combinations thereof: post-traumatic stress disorder, PTSD, post-traumatic stress, exposure therapy, imaginal exposure, flooding, prolonged exposure, and *in vivo* exposure. Each search was limited to available English journals and books. The reference lists of identified references were then searched for any references that might have been overlooked during the initial searches.

2.3 Inclusion Criteria for Outcome Studies

Two studies, reported in the medical literature, indicated that treatment effects were inflated by between 30% and 41% in studies of poorer methodological quality (Moher et al., cited in Harvey et al., 2003). For this reason, the selection of treatment outcome studies for this overview was very stringent. Stepwise inclusion criteria included:

a) Only imaginal and *in vivo* exposure-based randomised-control and randomised-comparison trials, in which only specific, replicable, and manualised treatments were utilized, were considered. The research stresses these as the most important criteria for judging the validity of outcome studies

(Graziano & Raulin, 2000; Harvey et al., 2003). After the screening, 16 studies adhered to these inclusion criteria (Boudewyns & Hyer, 1990; Boudewyns et al., 1990; Bryant et al., 2003; Cooper & Clum, 1989; Echburúa, De Corral, Zubizarreta, & Sarasua, 1997; Foa et al., 1991, 1999; Keane et al., 1989; Marks et al., 1998; Paunovic & Öst, 2001; Pitman, Orr, Altman, & Longpre, 1996; Resick et al., 2002; Richards, Lovell, & Marks, 1994; Tarrier et al., 1999; Taylor et al., 2003; Thompson, Charlton, Kerry, Lee, & Turner, 1995).

b) Secondly, one of the trial's treatment conditions had to be "pure" exposure therapy (as defined on p. 5), in order to control for the influence that confounding treatment components might have on the comparison of results across studies. Six studies not meeting this criterion were then eliminated: Boudewyns and Hyer (1990), Boudewyns et al. (1990), Cooper and Clum (1989), Echburúa et al. (1997), Pitman et al. (1996), and Thompson et al. (1995).

c) Finally, only studies published after 1997 and which adhered to all the Gold Standards were included, as the Gold Standards were only published in 1997 by Foa and Meadows. Five studies remained for overview and evaluation in this assignment: Marks et al. (1998), Foa et al. (1999), Tarrier et al. (1999), Resick et al. (2002), and Taylor et al. (2003).

2.4 Evaluation of Treatment Outcome Studies

2.4.1 Treatment outcome methodology criteria

The methodology employed by many controlled treatment outcome studies is relatively poor (Harvey et al., 2003), rendering biased conclusions. Furthermore, the inconsistencies in the design and the reporting of trials have made the assessment of methodological quality and a comparison of outcome results across trials difficult (Sherman, 1998). It was therefore decided that the assessment of outcome studies for this assignment must be based on stringent, but validated, criteria. Consequently, other research methodology guidelines have been consulted, in addition to the Gold Standards.

2.4.1.1 APA task force

A Division 12 task force of the APA offered valuable guidelines on criteria for empirically validated treatment outcome studies. A template was provided for the construction and evaluation of clinical interventions in the area of behavioural health (Barlow & Hofmann, 1997). It requires that treatment outcome research be constructed on the basis of two simultaneous axes (see Addendum 1). The first

axis requires that guidelines take into consideration a given intervention's absolute and relative efficacy, and the quality of the studies on which the judgement is made, as well as their replicability. The second axis specifies guidelines that consider the applicability and feasibility of the intervention in the local setting where it is to be offered as well as the generalisation of an intervention of established efficacy.

2.4.1.2 The CONSORT statement

In the medical field, the Consolidated Standards of Reporting Trials (CONSORT) statement (Altman et al., cited in Harvey et al., 2003) was developed to address the inconsistencies in the reporting of trials, which made the assessment of methodological quality difficult. It provides a checklist (see Addendum 2) for reporting on a study's method and design, assignment and randomisation, masking (blinding), follow-up, analysis, reporting of results, and discussion of results.

2.4.1.3 The revised Gold Standards

In order to determine whether rigorous research methodology predicted more significant treatment effect sizes, Maxfield and Hyer (2002) quantitatively compared the Gold Standards to effect sizes from EMDR outcome studies. Based on their results, Maxfield and Hyer (2002) identified three additional methodological components, necessary for methodological rigour, not incorporated into the Gold Standards. These components include (1) ***no concurrent treatment***, and refers to concurrent treatments that can obscure true effects by diminishing construct validity, and may increase the amount of "noise" and the likelihood of a Type II error, (2) ***multimodal assessment***, which advocates the use of multimodal measures that can assess a wide range of pathology and outcome predictors via interview, behavioural, and physiological measures, and (3) ***adequate course of treatment***, a parameter that was created to cater for the relative unpredictable amount of sessions required, as number of sessions may vary from simple trauma (fewer sessions) to complex traumas (more sessions required). This final parameter allows for the flexibility with which a treatment program can be adjusted to suit the type of problem and/or for cases when too much treatment has been rendered to the participant.

2.4.1.4 Ethical practice

The ethical use of exposure therapy and the optimal method for applying it are currently debated (Foa, Zoellner, Feeny, Hembree, & Alvarez-Conrad, 2002; Harvey et al., 2003; Rothbaum & Schwartz, 2002). Researchers of exposure treatment for PTSD do not agree on the ethical implications of the inclusion of a wait-list control group into their research designs. One school of thought advocates the use

of wait-list control groups to control for the possibility of natural remission (Resick et al., 2002), while others deem the withholding of treatment for PTSD sufferers as unethical (Petersen, cited in Paunovic & Öst, 2001). However, no guidelines have been proposed regarding ethical questions of including wait-list control groups.

2.4.2 Evaluation process

Based on the Gold Standards (Foa & Meadows, 1997), the CONSORT statement (Altman et al., cited in Harvey et al., 2003), the revised Gold Standards (Maxfield & Hyer, 2002), and the suggestions for external validity criteria as proposed by the AHRQ (Barlow & Hofmann, 1997), a set of criteria for use in this assignment was developed and is shown in Table 2.1. In addition, a criterion for the ethical use of wait-list control groups was also included (Petersen, cited in Paunovic & Öst, 2001). The 21 criteria are grouped into seven groups. The five identified treatment outcome studies will be assessed in terms of whether or not it has met these 21 criteria, either in full, partially, or not at all. The main methodological limitations (when three or more of the outcome studies share a specific limitation) will then be identified and discussed.

Table 2.1

Assessment Criteria for Treatment Outcome Studies (adapted from Barlow & Hofmann, 1997; Foa & Meadows, 1997; Altman et al., cited in Harvey et al., 2003; Maxfield & Hyer, 2002)

Criterion group	Criterion number	Criterion evaluated	Description of criterion
Motivation	1a	Rationale	Does the study have a scientific explanation of the rationale?
	1b	Objectives	Does the study have specific hypotheses?
Participants	2a	Exclusion criteria	Clear description of exclusion and inclusion criteria, including co-morbid conditions that must be excluded.
	2b	Target symptoms	This criterion equates to parameter one of the Gold Standards. Target symptoms must preferably extend further than Axis-I diagnosis, to include core symptoms of the relevant Axis-I diagnosis (e.g., for PTSD, symptoms of re-experiencing, intrusions, avoidance and numbing, and arousal must be specified as target symptoms).
Measures	3a	Reliable and valid measures	This criterion relates to parameter two of the Gold Standards, and includes specification of primary and secondary outcome measures, whether a multimodal approach is taken, and if large batteries are avoided (Harvey et al., 2003).
	3b	Assessor training	This criterion relates to parameter four of the Gold Standards, with the addition of therapist training.
	3c	Blindness	This criterion relates to parameter three of the Gold Standards, and includes participant blindness to treatment allocation, and how blindness is assessed.
Treatment	4a	Manualised, replicable, specific treatments	This criterion relates to parameter five of the Gold Standards, and includes detail regarding how and when the treatments were administered.
	4b	Sample size	How was sample size determined? To achieve a Power of at least 80%, a minimum sample size of 15 is considered good (see Dallal et al., in Marks et al., 1998). The study must further state differences in sample size at the various points of assessment, including that of treatment completers and dropouts.
	4c	Treatment adherence ratings.	This criterion relates to parameter seven of the Gold Standards, and includes adherence to and integrity of homework assignments, as well as estimates of treatment integrity and therapist characteristics.

Criterion group	Criterion number	Criterion evaluated	Description of criterion
Design and procedures	5a	Randomisation process	This criterion relates to parameter six of the Gold Standards, and includes what methods were used to carry out the allocation process, implement allocation, and ensure blindness?
	5b	Implementation of design	Are the procedures of the research design described in detail? This must include participant flow, dropouts, and deviations from the study as planned.
	5c	Statistical methods	The motivation for choice of statistical analyses and definite inclusion of relevant inferential statistics.
	5d	Ethical practice	Do studies have a wait-list control group, or comparison group? Are control group participants offered treatment? If so, is it via randomisation or of choice? Do studies address participant education regarding treatment disadvantages?
Results	6a	Numbers analysed	Does statistical analyses focus on both the intent-to-treat (ITT) and completer samples?
	6b	Baseline data	Does the study offer pre-treatment baseline demographic and clinical characteristics, including co-morbidity, of each treatment group?
	6c	Outcomes	A summary of results for all primary and secondary measures must be given.
	6d	Clinical improvement	This includes the reporting of effect sizes, end-state functioning, participant ratings of social functioning, diagnosis remission, and clinical improvement.
Discussion	7a	Interpretation	Interpretation of results, taking into account study hypotheses, study limitations, and sources of potential bias.
	7b	Overall evidence	General interpretation of the results in the context of current evidence.
	7c	Generalisation	Does the study have external validity with regard to patient feasibility, socio-cultural generalisation, and cost-effectiveness?

CHAPTER 3
OVERVIEW AND EVALUATION OF TREATMENT OUTCOME STUDIES FOR EXPOSURE TREATMENT OF PTSD

3.1 Introduction

This chapter provides a qualitative overview of the five identified exposure-based treatment outcome studies on PTSD, evaluate these studies according to the criteria described in Table 2.1, and identifies significant methodological limitations that might restrict the validity of the treatment outcomes.

3.2 Overview of Outcome Studies

3.2.1 Marks et al. (1998)

Marks et al. (1998) conducted a randomised control trial to investigate the efficacy of prolonged exposure (PE), cognitive restructuring (CR), and PE and CR combined (ECR), compared to relaxation (R).

Participants

Participants were outpatients referred between 1992 and 1995 by professionals, police and through support programs, or who presented themselves. They were of diverse trauma origin and mixed gender. Exclusion criteria were (a) ages younger than 16 years or older than 65 years, (b) suicidal intent, (c) organic brain disorders, (d) past or present psychosis, (e) antidepressant medication beyond that of a specified protocol, (f) alcohol misuse, and (g) past exposure treatment or cognitive treatment for PTSD. Two "blind" clinicians, via a two-phase interview method, made the diagnosis of PTSD according to DSM-III-R criteria. The diagnosis of PTSD had to be present for six or more months. None of the treatment groups differed from one another on demographic variables.

One hundred and nine participants met the inclusion criteria and were offered treatment. Of these, 22 refused and 87 began treatment (23 in PE, 19 in CR, 24 in ECR, and 21 in R). A further 10 dropped out before post-treatment assessment, and 77 (20, 18, 19, and 20 in the PE, CR, ECR and R groups respectively) completed 10 treatment sessions during a mean of 16 weeks. Of the 77 treatment completers, 52 (13, 12, 13, and 14 in the PE, CR, ECR and R groups respectively) completed follow-up assessment at week 36.

Measures

Assessment consisted of 12 primary measures, including (a) the Clinician-Administered PTSD Scale (CAPS-2) (assessor-rated), total symptoms score and severity, (b) the Impact of Events Scale (IES), (c) the PTSD Symptoms Scale (PSS), (d) the Beck Depression Inventory (BDI), (e) the State-Trait Anxiety Inventory (STAI), (f) the Fear Questionnaire (FQ), (g) the General Health Questionnaire (GHQ), (h) Global Improvement (GI) (assessor- and self-rated), (i) total of four goals (assessor- and self-rated), (j) Problem (assessor- and self-rated), and (k) Work/Social adjustment (assessor- and self-rated). Secondary measures consisted of 22 scales, mainly subscales of primary measures.

Treatments

Participants were randomised to one of four treatment conditions: (a) PE ($n = 23$), (b) CR ($n = 19$), (c) ECR ($n = 24$), and (d) R ($n = 21$). The latter condition was included as a placebo control for therapist contact and for homework practice between sessions. Two trained and experienced therapists (Lovell and Thrasher) conducted the treatments, using a procedure manual and four treatment manuals, covering each session in each treatment condition. Participants had ten individual 90-minute sessions, except for ECR, where sessions lasted 105 minutes. Sessions were audiotaped and rated. Cell sizes for the treatment conditions, based on a minimum power of 80%, were estimated at a minimum of 13.

The PE consisted of imaginal exposure in sessions one to five, using the audiotaped sessions as homework, rating peak distress before sessions, at critical points during sessions, at the end of sessions, and during homework in a daily diary. Sessions six through ten involved prolonged *in vivo* exposure with the aid of a SUD scale. Between-session homework entailed live exposure for one hour daily, recorded with distress levels in diaries.

Cognitive restructuring involved the protocols suggested by Beck and colleagues (see Beck, Emery, and Green, cited in Marks et al., 1998), excluding any exposure-based behavioural exercises.

Sessions one through five of ECR each involved 45 minutes of imaginal exposure, a 15 minute break, then 45 minutes of CR. In sessions six through ten, the imaginal exposure was substituted with *in vivo* exposure. Daily homework assignments of exposure-based exercises and cognitive tasks were encouraged, with daily recording of such assignments in diaries.

Progressive muscle relaxation, as proposed by Jacobsen (cited in Marks et al., 1998), was taught to participants in the R group. Participants were asked to do an hour's relaxation homework daily and to record anxiety during the homework in a diary.

Procedures

Referred participants were sent a screening questionnaire. Suitable respondents had a two-hour screening interview covering diagnosis, the trauma and its aftermath, clinical features, impact, mental status, and the Structured Clinical Interview for DSM-III-R for PTSD, other anxiety disorders, and substance abuse. Suitable participants were rated in a second two-hour interview to confirm adherence to the inclusion criteria and were then randomly assigned in permuted blocks of 20 to one of the four treatments, stratified for personal (intended by someone) or impersonal (e.g., accidents) trauma. The therapists were then randomly allocated to the groups.

Following pre-treatment assessment (week 0) by one of two blind assessors, treatment commenced, where after assessment occurred at weeks 6, 11 (post-treatment), and at 1-, 3- and 6-month follow-up.

Nonparametric analyses were used for categorical variables, while ANCOVA's and ANOVA's were used for continuous variables. All post-hoc comparisons were computed using least-significant difference scores (LSD), supported by Scheffé pairwise comparisons. For a few comparisons, t-tests were done.

Results

Within PE, CR, and ECR, improvement from pre-treatment to post-treatment and from pre-treatment to follow-up was highly significant on most primary measures. Compared to R, the PE, CR, and ECR groups exhibited significant improvement on 11 out of the 12 primary measures at post-treatment level, six out of 12 at 1-month follow-up, and eight out of 12 at 3-month follow-up. For end-point imputation analyses on the 12 primary measures PE, CR, and ECR improved significantly more than R from pre-treatment to 1-month follow-up on seven measures and from pre-treatment to 3-month follow-up on five measures. For pre-treatment to 6-month follow-up scores, E was significantly better than CR on eight primary and six secondary measures, but none was significant on end-point imputation analyses of primary measures. When change was computed for weeks 0 to 11, there was almost complete overlap between the confidence intervals of PE, CR, and ECR.

The effect size[1] estimates for PE, CR, and ECR was 1 to 2.5 from week 11 onward on most primary measures, and higher still on some primary measures for PE and ECR at 3- and 6-month follow-up. Ef-

[1] Effect size was computed as the mean change since week 0 divided by the standard deviation of that change, where ≥ 1.0 was regarded as clinically meaningful.

fect sizes in the R group were almost always smaller than in PE, CR, and ECR but were often 1.0 or larger. Clinical significance for each treatment condition was calculated for a few primary measures at post-treatment level. Significant percentage improvement on the IES was 60% in PE, 50% in CR, 58% in ECR, and 20% in R, and for the CAPS, 47% to 53% in the PE, CR, and ECR and 15% in the R.

Percentage of participants improved[2] was analysed for the IES, CAPS, and for GI. The PE, CR, and ECR groups consistently improved similarly, and more than the R group. Percentage improvement at post-treatment assessment was as follows: on the IES, 60% for PE, 50% for CR, 58% for ECR, and 20% for R, and on the CAPS, 47% to 53% for the PE, CR, and ECR groups, and 15% for R.

Of the 87 trial entrants, 74 (85%) consented to rating of audiotaped sessions by a blind assessor. Randomly selected tapes were rated for 17% of PE (34 tapes), 15% of CR (28 tapes), 13% of ECR (25 tapes), and 8% of R (17 tapes). Integrity ratings were satisfactory: on a scale of 0 to 8 (8 = excellent), mean ratings were similar across treatment conditions and therapists, for non-specific factors as well as for specific factors.

Discussion

Marks and colleagues evaluated the design of their study against the Gold Standards, and in addition, provided three more reasons to lend validity to the study's results: (a) assessors could not guess treatment assignment, (b) outcome was similar regardless of various variables, and (c) measured improvement was comparable to the results of previous studies. They also outlined three main limitations of the study: (a) despite randomisation, PE and R yielded lower scores on some baseline measures than CR and ECR, (b) in PE and ECR, non-completers at follow-up experienced symptom worsening since the last assessment, more so than completers at the same point in time, resulting in the slight advantage of PE over CR at 6-month follow-up being discarded, and (c) the trial used a multi-measure approach, thus increasing the chance of significant differences appearing randomly across groups.

On main outcomes, Marks et al. (1998) stated that, compared to previous randomised-control trials and randomised-comparison trials, current improvement on symptom severity and comorbid depression "...were at least as great." (p. 323). The authors accurately reported the equivocal outcomes of PE, CR, and ECR, but concluded that PE and CR were slightly superior. Due to the restricted nature of the

---

[2] For the IES and CAPS it was a change of 2 standard deviations or more improvement since week 0; and for the GI a criterion of markedly, much, or very much improved (Marks et al., 1998).

format of reporting of results, specific opinions regarding specific symptoms and dependant variables were not discussed. Marks et al. (1998) recommended the possible synthesis of exposure therapy and cognitive restructuring in future outcome studies.

Evaluation

Eighteen criteria were met in full (1a, 2a, 3, 4, 5, 6, 7bc), and three partially (1b, 2b, 7a). Although Marks et al. (1998) provided an appropriate rationale for their study, no experimental hypotheses were proposed (1b). Despite a wide range of outcome measures, no analyses of core symptoms (2b) were provided, only analyses of total scores. Due to the lack of research hypotheses, the interpretation of results did appear without context at times (7a). In addition, self-rated social improvement was lower than assessor rated improvement, indicating that participants did not perceive personal gain from the treatment. At post-treatment, 47% (n = 9) of PE, 68% (n = 12) of CR, and 68% (n = 12) of the ECR participants did not show good end-state functioning.

Marks et al. (1998) presented a well-designed trial, and reported it systematically. The study concluded that PE and CR showed similar results in terms of clinical improvement, and were superior to ECR. Limitations of the study included the following. *Firstly*, no inferential statistics on the core symptoms of PTSD were provided. This reduces the different treatment effects found between the treatment conditions. *Secondly*, the baseline differences of dropouts at pre-treatment, and their level of functioning at the time they terminated treatment, discounted the slight superiority of E over C at 6-month follow-up. *Thirdly*, the large assessment battery implemented, both self- and assessor-rated, could have increased the chances of significant differences appearing randomly across groups.

In conclusion, analyses of the Marks et al. (1998) data indicate a comparable efficacy for PE, CR, and ECR. The disappointing results for clinical improvement do however put the actual effectiveness of the active treatments in doubt. Irrespective of the results' valence, the heterogeneous sample (population and trauma type), made generalisation of results to the general population more reliable.

3.2.2 Foa et al. (1999)

Rationale

Citing major research outcomes to date, Foa et al. (1999) investigated the therapeutic effects of a combination of exposure therapy and Stress Inoculation Training (SIT) in reducing PTSD in female assault victims. Participants were randomly assigned to one of three conditions: SIT, prolonged exposure (PE), or a combination (PE-SIT), and a wait-list (WL) group. Foa and colleagues hypothesised that all

active treatments would be superior to the WL condition, and that the PE-SIT group would be superior to PE and SIT separately.

Participants

Participants were 96 female victims of assault (sexual or nonsexual) who met the diagnosis of PTSD according to DSM-III-R criteria. Exclusion criteria were current schizophrenia, bipolar disorder, organic mental disorder, alcohol or drug dependence, severe suicidal ideation, or being in an ongoing intimate relationship with one's assailant. Treatment groups did not differ significantly on demographic variables and pre-treatment measures of psychopathology. Twenty-one women refused treatment, 17 participants dropped out of treatment, leaving 79 completers. Dropout rate differed significantly across groups, with more participants dropping out from the SIT and PE-SIT groups. At 12-month follow-up, data was available for only 46 participants.

Measures

Two interview measures were used: (a) the SCID, and (b) the PSS. Foa and colleagues also utilized the Global scale of the Social Adjustment Scale (SAS) to assess functioning in various social and occupational settings. Self-report measures included the BDI, and the STAI's State Anxiety scale. Assessments were conducted at pre-treatment, post-treatment, and at follow-up at 3- ($N = 56$), 6- ($N = 54$), and 12-months ($N = 46$), with various blind female assessors rating treatment outcome. All measures except for the SCID were administered at each assessment point. Possible therapist effects were examined by means of an ANCOVA on the PSS-I, adjusting for pre-treatment severity. Blind assessors via videotaped treatment ratings of 52 treatment components determined treatment adherence.

Procedures

Participants were screened by telephone, with a subsequent personal screening, which included the SCID and the PSS-I. They were then randomly assigned to one of the four treatment groups. After enrolling 10 participants into the WL group, more were assigned to the three active treatment groups. After the pre-treatment assessment, treatment commenced.

Treatments

Participants were randomly assigned to one of three treatment conditions: (a) PE ($n = 23$), (b) SIT ($n = 19$), (c) PE + SIT ($n = 22$), and (d) WL ($n = 15$). Seven PhD-level clinical psychologists, trained to use a specified treatment manual, conducted treatment. These clinicians were under the ongoing supervision of Foa and Dancu. Treatment consisted of nine twice-weekly sessions, two initial sessions of 120

minutes followed by seven sessions of 90 minutes each. Following a five-week period from pre-treatment assessment, the WL participants were offered treatment.

Prolonged exposure consisted of imaginal and in vivo exposure, with sessions three to nine devoted explicitly to exposure activities for up to 45 – 60 minutes, using a SUD scale, including homework assignments of both an imaginal and *in vivo* nature. The protocol for PE was adapted from Foa and Meadows (cited in Foa et al., 1999). SIT was based on the methodology applied by Foa et al. (1991). PE-SIT treatment followed the nine-session format and included education, training in all the SIT skills, *in vivo* exposure, and imaginal exposure. Each session was conducted in the following order: brief homework review, imaginal exposure for 30-45 minutes, training in one coping skill, and homework consisting of both exposure and coping skill practice. WL participants were contacted once between assessments to ascertain the need for specialized intervention.

Results

Preliminary results indicated no significant therapist effects, nor significant differences in dropout rates. Videotapes of 63 therapy sessions (9% of the sessions) were randomly selected and rated. On average, therapists completed 93% of the treatment components.

For scores on the PSS, BDI, and STAI-S, the PE, SIT, and PE-SIT groups yielded superior outcomes compared to the WL group at post-treatment, for both the ITT and completer samples. Follow-up analyses also showed significant effects on all three measures, for both the ITT (PSS, BDI, and STAI-S) and completer (PSS, BDI, and STAI-S) samples. Simple comparisons for the ITT sample indicated that the PE group scored lower than the WL group on all three measures, whereas the SIT and the PE-SIT groups scored lower than WL on the PSS and BDI. For the completer sample, simple comparisons indicated that all three active treatment groups scored significantly lower than WL participants on all three outcome measures (PSS, BDI, and STAI).

End-state functioning was defined as a cut-off score of 20 on the PSS-I, 40 on the STAI-S, and 10 for the BDI. At post-treatment level revealed similar results for both the ITT and the completer samples, with 52% and 57% of the PE group respectively, 31% and 42% of the SIT group, and 27% and 36% of the PE-SIT participants achieving satisfactory functioning, as compared to 0% of the WL completers. Analyses for diagnostic status at post-treatment showed that 60% and 65% of the PE group, 42% and 58% of the SIT group, and 40% and 54% of the PE-SIT group, ITT and completer participants respectively, lost their diagnosis. These are significant decreases relative to the WL participants, but not between the groups themselves.

Cohen's *d* statistics were calculated for the effect sizes at post-treatment. The PE group yielded the largest effect sizes for all three outcome measures compared to the SIT and PE-SIT conditions, for both the ITT and completer samples (1.46, 1.42, and 1.32 for the PSS, BDI, and STAI respectively for the ITT sample, and 1.92, 1.47, and 1.44 for the PSS, BDI, and STAI respectively for the completer sample).

At follow-up, simple effects analyses revealed that PE completers (n = 16 at 12-months follow-up) scored significantly lower on the STAI than did SIT (n = 14) and PE-SIT (n = 16) completers. PTSD diagnosis ranged between 65% and 68% at 12-months follow-up. On the SAS, the PE group scored significantly lower than the WL group. No differences were detected between the three active treatments at follow-up.

Discussion

Foa and colleagues conceded that the results did not support the hypothesis that PE-SIT would be superior to the PE and SIT conditions, but provided no plausible explanation for the apparent supremacy of PE, except to note the lower dropout rates for PE. Instead, various factors were suggested for the "unexpected" (p. 199) failure of PE-SIT. Drawing on results by Marks et al. (1998), Foa et al. (1999) suggested that a combination of PE and SIT might be too strenuous on participants. Two methodological limitations that might have hampered the success of SIT were noted. The discussion was concluded with the endorsement of PE as an effective treatment due to the ease with which it can be administered by non-expert clinicians.

Evaluation

Fifteen of the evaluation criteria (1, 3, 4ab, 5bc, 6, 7ab) were complied with in full, and six (2, 4c, 5ad, 7c) partially. No mention was made of the method of recruitment (2a) and referral. As in the study by Marks et al. (1998), Foa et al. (1999) did not investigate the core symptoms of PTSD (2b), only total scores of outcome measures. Treatment adherence was estimated, but not homework adherence (4c). Although participants were randomly assigned, no mention was made of the randomisation process (5a). A wait-list control group was included (5d), and offered treatment after post-treatment assessments. Whether or not these WL participants had a choice in the type of treatment received was not mentioned. Results of the study are only partially generalisable (7c), due to the specific population and trauma type under investigation.

Foa and colleagues concluded that PE was marginally superior to SIT and PE-SIT, and that SIT and PE-SIT were of equal efficacy. Important sources of bias need to be taken into account. *Firstly*, symp-

tom change was not analysed in enough detail. Lovell, Marks, Noshirvani, Thrasher, and Livanou (2001) showed that additional data analyses changed previous treatment outcome results by Marks et al. (1998). *Secondly*, no information is available on the level of functioning of dropouts during the active treatment phase. Because significantly more dropouts occurred for SIT and SIT-PE during this time, results could have underestimated the impact of these treatment conditions (according to ITT sample analysis). *Thirdly*, only 48% ($n = 46$) of the ITT sample remained for follow-up analyses. Although the follow-up sample sizes for groups were still big enough, the considerable decrease in sample size from pre-treatment does decrease the reliability of results.

Given the crucial impact that the sources of bias have on the interpretation of results, it is suggested that the marginal superiority of PE be ignored, and that it be equated to the efficacy of SIT and PE-SIT. In addition, results from the clinical improvement of participants are important. At 12-months follow-up, 48% ($n = 7$) in PE, 58% ($n = 8$) in SIT, and 64% ($n = 10$) in PE-SIT did not achieve good end-state functioning. For all three treatment groups, approximately 33% of participants still met PTSD diagnosis at 12-months follow-up. Thus, the overall benefits obtained by participants are in doubt.

3.2.3 Tarrier et al. (1999)

Rationale

According to Tarrier et al. (1999), confounding components in treatments make direct comparisons difficult. Tarrier et al. (1999) addressed the afore-mentioned problem by comparing imaginal exposure (IE) to cognitive restructuring (CR), having removed all exposure elements from the CR treatment and discussion of thoughts and emotions from the IE treatment condition. The authors advanced two competing hypotheses: (a) directly changing dysfunctional beliefs and cognitions resulting from the trauma is essential, and (b) therapeutic benefit will only result from direct exposure to trauma memories.

Participants

Primary and secondary health services referred 204 patients of mixed gender and trauma type for screening as possible participants. PTSD diagnosis was based on DSM-IV criteria. Exclusion criteria included current alcohol and/or substance abuse, psychotic or organic brain disorders, childhood sexual abuse as the index trauma, inconsistent medication during and up to three months prior to participation in the study, and receiving any concurrent psychological intervention in the six months preceding the referral. After three phases of pre-treatment assessment, 72 participants were allocated to either an IE ($n = 35$) or CR ($n = 37$) group, of which 62 completed treatment (IE, $n = 29$; CR, $n = 33$), and 57 (IE, n

= 27; CR, n = 29) were available for 6-month follow-up assessment. CT participants only differed from IE participants in that a significant number had a previous psychiatric disorder. Thirty-three (54%) participants had current major depression, with a further 11 (18%) having past major depression. Twenty-six (43%) were diagnosed with general anxiety disorder, 15 (25%) with panic disorder, and 6 (10%) had a simple phobia.

Measures

To assess treatment expectations and impressions, participants completed the Credibility of Treatment Questionnaire (CTQ) and were rated by the therapist for treatment motivation on a 6-item scale. Treatment was followed by post-treatment assessment and six-month follow-up assessment by a "blind" assessor. Additional logistic measures[3] were implemented to ensure treatment "blindness" to assessors and participants. The following assessment measures were conducted at each assessment point: (a) the CAPS for symptom change and symptom severity, (b) the Penn Inventory (PI) for symptom severity, (c) the IES for indications of intrusion and avoidance symptoms, (d) the BDI, (e) the Beck Anxiety Inventory (BAI), and the General Health Questionnaire 28 (GHQ-28) for an indication of general psychopathology. Two "blind" assessors rated 8.4% of the sessions for therapist adherence to the treatment manuals.

Treatments

Treatment consisted of 16 one-hour sessions and was conducted by two of the authors (Tarrier and Pilgrim) who frequently met for peer supervision. The CT manual was developed using the principles and techniques suggested by Beck and Emery (cited in Tarrier et al., 1999) and Resick and Schnicke (cited in Tarrier et al., 1999). Discussion of the trauma incident itself was avoided, instead therapy aimed at eliciting participants' beliefs about the meaning of the event and attributions made following it. The IE manual was adopted from Foa et al. (1991), and focused on exposure to the trauma memory.

Procedure

Assignment to treatment entailed a three-phase assessment process, beginning with an initial interview and screening upon referral (phase I), followed by a fuller assessment based upon DSM-IV criteria and the CAPS (phase II). Following phase II was a four-week baseline self-monitoring period, after which re-assessment by the CAPS (phase III) ensured admission into the study. A "blind" clinician, using a

---

[3] These measures included the use of separate administrative procedures between therapists and assessors, multiple coding of

(Footnote continues on next page)

process of minimization, stratified on various demographic variables, did a software-based random allocation of participants to the treatment conditions. During session two, participants completed the CTQ and at the end of session five therapists rated the participants' motivation for therapy.

Results

Treatment manual adherence ratings (double rated) of 8.4% of the sessions (37 audiotapes, 7 videotapes) indicated excellent interrater reliability ($\kappa = .947$). Dropout rates, treatment credibility, expectancy of benefit, and participant motivation did not differ between the treatment groups.

Repeated measures ANOVA's indicated that for both IE and CR, all measures revealed a significant decrease from pre-treatment to follow-up, indicating symptom improvement. There were no significant differences between E and CR on any of the outcome measures. Individual analyses of IE and CR indicated that for IES (avoidance symptoms), IES (intrusions) and GHQ-28, changes over time did not reach significant levels.

Pre-post effect sizes indicated larger effect sizes for CR on all measures. Assessor ratings (CAPS Global Severity scale, GSS), compared with self-ratings (CAPS Global Improvement scale, GIS), were used to calculate clinical significant improvement. Both these measures showed no significant differences between the treatment groups: for GSS, post-treatment IE = 41%, and CR = 33%, and follow-up IE = 26%, and CR = 35%; for GIS, post-treatment IE = 48%, CR = 27%, follow-up IE = 26%, and CR = 17%. At post-treatment level, IE showed non-significant superiority regarding the number of participants maintaining the PTSD diagnosis (E = 41%, CT = 58%), but this difference disappeared at follow-up level (both E and CT = 48%).

Analysis of treatment failure indicated that significant more IE participants (IE = 9, CR = 3) worsened on the total CAPS score between pre- and post-treatment. It was found that these non-improvers, compared to the 50 improvers, rated therapy of significantly less value, were assessor rated as significantly less motivated, and missed a significantly greater number of therapy sessions.

Discussion

Tarrier and colleagues conceded that neither of the two hypotheses posed were proved, but added that within the realm of anxiety disorders, the failure of various cognitive-behavioural approaches to dem-

---

treatment allocation, independent data management, and instruction to patients not to discuss their treatment with the assessor.

onstrate superiority was not uncommon. Apart from treatment failure causally linked to non-compliance, no plausible explanations were given for non-compliance. The authors acknowledged several limitations of the study, including: (a) the absence of a wait-list control group to control for spontaneous remission, (b) the IE and CR used differed slightly from other similar manualised forms of treatment, and (c) treatment fidelity was not assessed in detail. The authors concluded that the results of the study are "encouraging" (p. 17).

Evaluation

Eighteen of the evaluation criteria (1, 2, 3ac, 4ab, 5, 6bcd, 7) were met in full, three partially (3b, 4c, 6a). No information is provided on assessor training (3b), and no adherence to homework was assessed (4c). In addition, criteria used to assess treatment adherence is unclear. Statistical analyses of data only applied to the completer sample (6a).

Tarrier and colleagues concluded that neither imaginal exposure nor cognitive restructuring was superior to one another, nor that neither treatment condition was sufficient for complete symptom reduction. Another source of bias was the actual time frame in which treatment transpired. Treatment was planned to take place over 16 weekly sessions, but the average attendance was once every two weeks, for this could have reduced the accumulating effect of treatment. Clinical improvement paints a gloomy picture: participant-rated end-state functioning yielded good results for only 26% (n = 7) of the IE group, and 17% (n = 5) of the CR group at follow-up. At 6-months follow-up, 48% of IE (n = 13) and CR (n = 14) remained PTSD cases. Finally, both treatment conditions differed slightly from other similar manualised forms of treatment, raising doubts concerning the generalisability of results.

From the mentioned sources of bias, it would appear that the Tarrier et al. (1999) study is limited in two respects. *Firstly*, the limited generalisability due to the deviation from accepted treatment protocols and *secondly*, results showed limited clinical improvement.

3.2.4 *Resick et al. (2002)*

Rationale

Resick et al. (2002) stated that no controlled study exists which compare cognitive processing therapy (CPT) with other validated treatment options for PTSD-diagnosed rape victims. In order to address this shortcoming, a randomised control trial of CPT compared to prolonged exposure (PE) and a minimal attention wait-list condition (MA) was designed. A secondary purpose of the study was to examine the effect of both therapies on dysfunctional cognitions, specifically self-blame and guilt. It was hypothe-

sised that CPT would be more effective than PE in altering guilt cognitions.

Participants

Two hundred and sixty-seven female rape survivors were assessed for possible inclusion in the study, with a diagnosis based on DSM-IV criteria. Exclusion criteria included current psychosis, unstable functioning while on current medication, less than three months post-trauma, developmental disabilities, suicidal intent and current para-suicidal behaviour, current dependence on drugs or alcohol, illiteracy, current involvement in an abusive relationship, incest as the index rape, and substance abuse within six months prior to the study. Of 181 women accepted into the trial, 10 were excluded due to emerging exclusion criteria. Of the remaining 171 women (CPT = 62; PE = 62; MA = 47), 13 never returned and 37 dropped out. One hundred and twenty-one women completed treatment (CPT = 41, PE = 40, MA = 40). Dropout rates did not differ significantly between CPT and PE. There were no significant differences between the treatment groups regarding demographic variables. Approximately 85% of the sample had experienced at least one other major crime victimization in addition to the index rape. At 3-months follow-up, participant numbers dropped to 74 (CPT = 37; PE = 37), and at 9-month follow-up to 52 (CPT = 26; PE = 26).

Measures

Selection for participation was done via individual interviews that included the CAPS, the Structured Interview for DSM-IV-Patient Version (SCID), and the standardised trauma interview (STI) which included the Sexual Abuse Exposure Questionnaire (SAEQ). Self-report measures included the PSS, the BDI, and the Trauma-Related Guilt Inventory (TRGI). In order to assess the expectancy of therapeutic outcome, Resick and colleagues administered a qualitative questionnaire designed by Foa et al. (1991) at pre- and post-treatment. "Blind" evaluators assessed treatment adherence and therapist competence. All sessions were videotaped and available for random assessment, which focused on sections on unique and essential elements specific to each session, essential but not unique elements, acceptable but not necessary elements, and prescribed elements for each therapy (Nishith & Resick, cited in Resick et al., 2002).

Treatments

Resick et al. (2002) randomly assigned participants to one of three treatment conditions: PE (n = 40), CPT (n = 41), or MA (n = 40). Eight female therapists conducted therapy, with doctoral training in cognitive-behaviour therapy and additional training in CPT and PE. Continuous supervision was provided throughout the study. Each therapist handled approximately an equal number of cases in each

condition. The two active treatment conditions comprised of 90 minutes sessions, for a total of 13 hours of treatment, spanning six weeks. Both PE and CPT involved considerable homework assignments, but due to treatment protocols, the two treatments' homework loads could not be equated, resulting in PE participants having significantly more homework.

Prolonged exposure was conducted according to the protocol developed by Foa and Rothbaum (cited in Resick et al., 2002) and Foa et al. (cited in Resick et al., 2002), and included psycho-education, breathing retraining, and imaginal and *in vivo* exposure. Homework involved listening to tapes of the imaginal sessions and engaging in *in vivo* behavioural exercises. Cognitive processing therapy followed the manual by Resick and Schnicke (cited in Resick et al., 2002), with only minor modifications, and included psycho-education, written narratives on several topics, and cognitive therapy based on emergent themes from the narratives. The MA condition served as a wait-list control group. Participants in MA were ensured of therapy in six weeks time, with diagnostic check-ups every two weeks to rule out emergency intervention. Various measures were put in place to ensure that MA participants had access to immediate therapeutic intervention if they required it.

Procedures

Screening of participants was followed by random allocation of participants to one of the two treatment conditions, after which pre-treatment assessment took place. Interrater reliability of the interview and screening process was established by having videotapes of the sessions rated independently by senior faculty members. Interrater reliability reached Kappa (κ) values in excess of .69 for all symptom criteria measured. Treatment followed, with the two active treatments completed in six weeks. At the end of the post-treatment assessment, MA participants were randomly assigned to either CPT or PE. Follow-up assessments were done at 3- and 9-months after post-treatment.

Results

Interrater reliability of the interview and screening process, treatment adherence, and therapist competence were all judged to be satisfactory in every treatment condition. Therapy outcome expectancies indicated no significant differences between treatment completers and dropouts, or between the participants of the two active treatments.

Analyses of variance results from the ITT sample, using the "last observations carried forward" method indicated significant differences between the groups on the CAPS and BDI at post-treatment. A post-hoc Tukey test indicated that from pre-treatment to follow-up, both the CPT group and PE group scored significantly lower than the MA group on the CAPS and BDI. Multivariate ANOVA's for each

group across the four assessment periods indicated that both the CPT and PE groups changed significantly over time, but not from post-treatment to follow-up. Random regression analyses for CAPS and BDI scores confirmed the results obtained.

Effect sizes for the two active treatments were calculated using Hedges g effect sizes.[4] The CPT and PE groups showed larger effect sizes for symptom change, relative to the MA group. Comparing the CPT and PE groups directly, CPT resulted in small but positive effect-size differences for PTSD, depression, guilt and dysfunctional cognition measures at post-treatment, 3 months, and 9 months, indicating modestly larger symptomatic improvement than for the participants in the PE group.

Multivariate ANOVA's of treatment completers yielded significant interaction, treatment group, and session effects for CAPS and BDI as dependant variables. Additional ANOVA's indicated significant post-treatment effects on the CAPS and BDI. Post hoc Tukey tests indicated that the group differences on both the CAPS and BDI were between the MA group and the two treatment groups. On the CAPS, both groups exhibited a strong decrease in scores from pre-treatment to post-treatment, and little increase from post-treatment to the 3-month follow-up. For the PE group, CAPS scores decreased again significantly from 3- to 9-months follow-up, to a level prior that of post-treatment. On the BDI, the groups improved significantly from pre-treatment to post-treatment.

Both therapies had large effect sizes relative to MA at post-treatment on PTSD, depression, and guilt scores. Relative to PE, CPT exhibited similar effect sizes as what the ITT sample yielded. However, on the BDI, PE was moderately more superior.

At post-treatment, 47% of CPT and PE participants in the ITT sample still met PTSD diagnosis. At 3-month follow up it was 42% for CPT and 53% for PE, and at 9-month follow-up 45% for CPT and 50% for PE. For treatment completers, only 20% in the CPT group and 18% of the PE group still met criteria for PTSD at post-treatment. At the 3-month follow-up, 16% of CPT and 30% of PE participants was PTSD positive. At 9-month follow-up, 19% of CPT participants and 15% of PE participants met the criteria. In the completer sample, 76% of the CPT and 58% of the PE participants reported good end-state functioning[5]. At the 3-month follow-up, 72% of CPT and 50% of PE participants reported good end-state functioning. At 9-month follow-up, 64% of CPT and 68% of PE participants ex-

---

[4] According to Cohen (cited in Resick et al., 2002), small effect sizes are defined as 0.2; medium as 0.5; and large as 0.8.
[5] Good end-state functioning was defined as a cut-off score of 20 on the PSS and 10 on the BDI.

hibiting good end-state functioning.

Analyses of dysfunctional cognitions yielded similar results. For the ITT sample, groups were different on all four TRGI subscales (global guilt, hindsight bias; lack of justification; and wrongdoing) at post-treatment. Again, CPT and PE yielded better results than MA, with CPT showing a tentative superiority to PE. Identical results were obtained at 9-months follow-up, although no group improved significantly from post-treatment to follow-up. According to Resick et al. (2002), similar results were obtained for the completer sample, but none of the inferential statistics were supplied in the article itself.

Discussion

Resick and colleagues stated that both CPT and PE were superior to MA. The authors conceded that CPT was superior to PE in changing guilt cognitions on two of the four TRGI subscales, had better effect sizes, and showed a trend towards better end-state functioning. Taking into account that CPT utilised fewer sessions to deal directly with the trauma, the authors suggested that CPT was superior to PE in alleviating PTSD symptoms in female rape victims. It was concluded that hindsight bias and lack of justification were guilt cognitions that required more cognitive restructuring due to its idiosyncratic nature, hence CPT's apparent superiority to PE.

The researchers pointed out two methodological limitations: (a) the study did not investigate a sufficient number of variables that could have explained participant dropout, and (b) CPT was only tested with female rape victims. Resick et al. (2002) called for additional outcome research on the efficacy of CPT. The authors advocated research that would determine whether exposure and/or cognitive components are necessary and sufficient for alleviating PTSD symptoms. Finally, the authors recommended that "we need to begin matching clients with therapies for optimal outcomes" (p. 878), as treatment outcome studies indicate that more than one treatment modality might be effective for a particular condition.

Evaluation

Eighteen evaluation criteria were met in full, and three partially (5ad, 7c). Despite random treatment allocation, no information regarding the randomisation process (5a) was available. The study included a minimal attention wait-list group (5d), with great effort taken to ensure that participants in this group did not experience symptom worsening. The authors acknowledged the limited generalisability of their results (7c), as trauma type was restricted and cognitive-processing therapy has only been investigated with rape victims.

Resick and colleagues concluded that CPT is comparable to PE on treatment efficacy, but shows superiority in alleviating two cardinal dysfunctional guilt cognitions. Two aspects of the outcome measures must be considered. *Firstly*, no outcome measure for state anxiety was used, and *secondly*, specific symptoms investigated focused on dysfunctional cognitions and did not include PTSD core symptoms. Since exposure therapy focuses on alleviating state anxiety via addressing the PTSD core symptoms, the two factors mentioned might have biased results in favour of CPT by avoiding key assessment variables.

Resick and colleagues presented a generally well-designed and reported treatment trial, with few limitations in research design. Compared to previous studies, clinical improvement of participants reached more acceptable levels. Taking into account the possible bias towards CPT regarding choice of target symptoms and outcome measures, the slight superiority awarded to CPT is negated. Results are only generalisable to female rape victims.

3.2.5 Taylor et al. (2003)

Rationale

Taylor et al. (2003) conducted a comparative study to ascertain the true comparative efficacy, speed of symptom improvement, and adverse effects of three PTSD treatments, namely exposure therapy (E), EMDR, and relaxation training (RT), as few studies with well-controlled designs has attempted to incorporate EMDR into treatment.

Participants

Recruited from physician referrals and through local media advertisements, 164 participants were invited for screening. Sixty participants met the inclusion/exclusion criteria and 45 participants completed treatment. Exclusion criteria included mental retardation, current psychotic disorders, and commencement or change in dose of psychotropic medication within the past three months. The 60 participants that entered treatment represented a diverse population in terms of gender (78% = female), race (77% = Caucasian), education (78% = college education), socio-economic status, and multiplicity and diversity in the experienced trauma (sexual and physical assault and motor vehicle accidents most common). Comorbid conditions included major depression (42%), panic disorder (31%), and social anxiety disorder (12%). The ITT sample, allocated to a treatment condition, did not differ on any demographic variables.

Measures

Taylor et al. (2003) included primary and secondary[6] outcome measures. Structured interviews utilized were the SCID-IV for intake diagnosis on Axis I, and the CAPS for PTSD symptom severity. As part of the intake interview, an additional CAPS item (trauma-related guilt over acts of omission or commission) was included as a secondary outcome measure. For the post-treatment and follow-up CAPS interviews, questions were added to assess the occurrence of stressful life events and change in medication during treatment and follow-up. Self-report questionnaires to assess PTSD symptom severity included the PSS. For depression, the BDI was included and trauma related anger was measured by an item assessing the frequency of anger about the trauma events on a Lickert-type scale. Participants' perceived treatment credibility was measured via the Reactions to Treatment Questionnaire (RTQ), administered at the beginning of session 2.

Treatments

Participants were randomised to eight 90 minutes individual sessions of E (n = 15), EMDR (n = 15), or RT (n = 15). Exposure and RT manuals were based on Marks et al. (1998), and EMDR on Shapiro (1999). Marks et al.'s protocol only differs from Foa's widely used Exposure protocol (Foa et al., 1999) in that it does not include breathing retraining. Exposure exercises occupied about 60 minutes of each 90 minutes session, consisting of repeated imaginal exposure to sensory-related disturbing associations, between-session exposure-related homework exercises, as well as therapist-assisted in vivo exposure for an hour each day for four weeks. Relaxation training involved three different relaxation exercises, which were practiced over three sessions. On selecting an exercise, participants practised it in subsequent sessions with the aid of a therapist-read relaxation script and between-sessions using audiotaped versions of the sessions. The EMDR sessions comprised of (a) practice in the Safe Place exercise, (b) recall of events and associated memories, (c) inducement of eye movement, (d) reporting of conjured associations, and (e) repetition of the process until provoked anxiety subsided. Two female therapists conducted the treatment protocols. Both worked under the supervision of one of the authors (Taylor), had several years of experience utilizing trauma-related therapeutic techniques, and had completed training from the EMDR Institute.

Procedure

Potential participants contacting the clinic were screened for inclusion-exclusion criteria during a tele-

---

[6] Secondary measures assessed commonly associated PTSD symptoms (guilt, anger, dissociation, and depression).

phone interview. Those successfully screened were invited for further assessment consisting of the SCID-IV, CAPS, and self-report questionnaires. Assessment of participants occurred at pre-treatment, post-treatment (one month after treatment ended) and at follow-up (four months after treatment ended). Blind evaluators conducted these assessments.

A "blind" doctoral-level psychologist, using a random selection of audiotaped sessions, assessed interrater reliability of the interview measures. Interrater agreement for the diagnosis of PTSD was 92% (κ = .80). Seven assessors, six of who were "blind", rated randomly selected videotaped sessions, comprising 59% of all treatment sessions. On all treatment integrity factors assessed[7], interrater agreement levels were above 86%.

Results

Analysis of results focused on treatment completers, although initial analysis indicated no significant differences between treatments for the ITT sample.

All three treatment conditions yielded significant reductions on the four dimensions of the CAPS scores, from pre-treatment to follow-up. Further ANCOVA's yielded only a significant time main effect for numbing. There were treatment main effects for re-experiencing and avoidance only. For re-experiencing and avoidance symptoms, E was significantly more effective than both RT and EMDR.

Clinical significant change, for each of the four dimensions measured by the CAPS, was defined as a reduction in scores of at least two standard deviations. At follow-up, E yielded significant better improvement over RT on re-experiencing, avoidance, and hyperarousal. Exposure yielded significant reductions in the percentage of participants still meeting criteria for PTSD, compared to RT, at each of the assessment periods.

Regarding secondary outcome measures, mean scores declined significantly from pre-treatment to follow-up for guilt, anger, and depression for each treatment. Dissociative symptoms significantly declined for the E and RT groups.

---

[7] Such factors included: (a) whether treatment-non-specific components such as therapist warmth and rapport were adequate; (b) whether treatment-specific components (e.g. imaginal exposure exercises) were implemented adequately; and (c) whether the session contained a non-protocol intervention, such as cognitive restructuring during E.

Discussion

Taylor and colleagues highlighted PE's relative superiority over both RT and EMDR. They further argued that "therapist skill" play a major part in the relative differences between treatment conditions. Two limitations were acknowledged: (a) a moderate sample size, and (b) the chronic nature of the sample that might limit generalisation to more moderate PTSD samples.

Evaluation

Nineteen criteria were complied with in full, and three partially (4c, 5a). Omitted from treatment adherence ratings (4c) was an estimation of homework adherence. Taylor et al. (2003) did not provide information on how randomisation occurred (5a).

Taylor et al. (2003) concluded that exposure therapy is a first-line psychosocial treatment for PTSD, with only limited efficacy awarded to EMDR. They presented a well-designed and reported treatment trial. Compared to the Resick et al. (2002) study, good remission from PTSD diagnosis for one of the treatments (exposure therapy) was obtained, with minimal symptom worsening for EMDR. However, no other measure was used to estimate significant clinical change. The biggest limitation of the study is generalisation. This was the first well-controlled study to compare the efficacy of EMDR to other psychosocial treatments, including exposure therapy. For this study, results are valid and generalisable to the general population (heterogeneous sample), but it is presumptuous to declare on the basis of one study that exposure therapy is the first-line of choice as psychosocial treatment for PTSD.

3.3 Summary

Chapter 3 provided an overview of five well-controlled treatment outcome studies of exposure therapy for PTSD (see Table 3.1, p.34, for a summary of the studies), and examined these studies' research methodology against 21 criteria. From this evaluation, the following methodological limitations have been identified:

a) Core symptoms of PTSD were not sufficiently assessed in three of the studies (Marks et al., 1998; Foa et al., 1999; Tarrier et al., 1999).

b) Detailed treatment adherence ratings were not provided. Estimates did not consistently include homework adherence and the influence of therapist variables. Further, as little as 8% of sessions were included for ratings (Foa et al., 1999; Marks et al., 1998; Tarrier et al., 1999), compared to

the 59% of Taylor et al. (2003).

c) Three of the treatment outcome studies (Foa et al., 1999; Marks et al., 1998; Tarrier et al., 1999) reported poor results in terms of significant clinical change. The "good" results reported by Taylor et al. (2003) were only based on estimates of symptom change. Resick et al. (2002) and Taylor et al. (2003) did not provide self-rated estimates of clinical change.

d) Statistical analyses of attrition data were insufficient. Tarrier et al. (1999) did not include analysis of attrition data at all, while the remaining four studies utilised the "last observation carried forward" (LOCF) method to analyse dropout data.

e) Two studies (Foa et al., 1999; Resick et al., 2002) utilised a wait-list control group in its research design. This raises doubts regarding the ethical use of exposure treatment in research trials.

Findings that emerged from the evaluations in this chapter must be validated against available literature before it can be considered significant. Chapter 4 will conclude the overview by discussing the limitations, where after the final conclusions and recommendations will be outlined.

Table 3.1

Summary of Exposure Outcome Studies for PTSD

Study	Trauma/ target population	Treatment conditions	*n*	Sessions	Results			
					Post-Test	Follow-up	Effect sizes	Between treatment conditions
Marks et al. (1998)	Mixed gender, various traumas	1. PE 2. CR 3. ECR 4. R	23 19 24 21	10	* 1,2, & 3 were superior to 4 on re-experiencing, avoidance/numbing, and associated features, but not on increased arousal. * No differences between 1, 2, and 3 in terms of improvement.	* 1-, 3-, and 6-mo maintained results from post-treatment. * 1, 2, and 3 equal * 2, 3 improved detachment	1, 2, and 3 ranged between 1 & 2.5, bigger than 4.	1 = 2 = 3 > 4
Foa et al. (1999)	Female victims of sexual and nonsexual assault	1. PE 2. SIT 3. SIT + PE 4. WL	23 19 22 15	9, twice weekly	* 1, 2, and 3 reduced symptoms, symptom severity, depression, and state anxiety more than 4 * 1 > 2 & 3 in reducing anxiety (ITT) * 1 > 3 in reducing depression *Good end-state functioning, reduction in diagnostic status for all three active treatments	At 12-mo: * Effects maintained * Anxiety significantly lower from post-treatment *PE greatest reduction in state anxiety	At post-treatment: PE greatest, ≥ 1.44 & ≤ 1.92	1 > 2 = 3 > 4, but 1 not always statistically significantly better
Taylor et al. (2003)	Mixed gender, diverse traumas, mainly sexual/physical assault & vehicle accidents	1. E 2. EMDR 3. R	15 15 15	8 x 90min each	* E reduced % that meet diagnostic criteria the most * E reduced re-experiencing, avoidance more than EMDR & RT * Numbing decreased the most of all 4 PTSD dimensions * E improved avoidance symptoms fastest	Results maintained	None	1 > 2 > 3

Study	Trauma/ target population	Treatment conditions	*n*	Results: Sessions	Post-Test	Follow-up	Effect sizes	Between treatment conditions
Tarrier et al. (1999)	Mixed gender, various traumas – mainly crime and accidents	1. IE 2. CR	29 33	16 x 60min weekly	* 1 & 2 improved all measures, except avoidance symptoms (IES) * 1 also did not improve intrusion symptoms (IES) and diagnosis remission.	* 6-mo: All results maintained * 1 & 2: 48% PTSD * 12-mo results maintained * intrusion and avoidance (IES) improved from 6-mo follow-up * 1 & 2: 38% PTSD	2 ≥ 1	1 = 2
Resick et al. (2002)	Female victims of diverse traumas, mainly rape	1. CPT 2. PE 3. MA	41 40 40	Bi-weekly – 13h total	*1, 2 > 3 in improving depression, overall PTSD symptoms, global guilt, wrongdoing cognitions. *1 had best results regarding hindsight bias, lack of justification cognitions *1 = 2 on end-state functioning and % participants that still has diagnosis	*Most results maintained *CAPS scores worsened from post-treatment to 3-mo follow-up. *2, CAPS scores lowered well from 3-mo to 9-mo follow-up	* Effect sizes in favour of 1 *2 showed advantage at 9-m follow-up for PTSD symptoms	1 ≥ 2

* CBT = Cognitive-behavioural Therapy; CR = Cognitive Restructuring; IE = Imaginal Exposure; IT = Implosive Therapy; MA = Minimal Attention control group; PE = Prolonged Exposure; R = Relaxation Training; ECR = PE + CR; IECR = IE + CR.

* "Between treatment conditions" indicate the relative efficacy as concluded by the study self.

CHAPTER 4
DISCUSSION AND RECOMMENDATIONS

4.1 Discussion of Main Findings

Chapter 3 identified several methodological limitations of the identified treatment outcome studies, despite their well-controlled status and their adherence to the Gold Standards. These limitations will now be discussed.

4.1.1 Assessment of core PTSD symptoms

Chapter 3 revealed a lack of assessment of core PTSD symptoms across the five treatment outcome studies. This relates very closely to the power[1] of a study. In any research design, the optimal balance would be to minimise Type I errors, and to maximise power to find real differences. The more detailed the target symptoms under investigation, the higher the power would be. However, three studies did not address the core symptoms of PTSD (Foa et al., 1999; Marks et al., 1998; Tarrier et al., 1999).

Foa and Meadows (1997) did not specify in parameter 1 of the Gold Standards that the core symptoms of a specific diagnostic target syndrome (e.g., PTSD) should be included in treatment outcome research hypotheses. However, not including changes in core symptoms may lead to an increase in Type I errors. This indicates a significant limitation in parameter 1 of the Gold Standards.

Establishing optimal power levels therefore becomes a matter of priority in terms of research design: what is essential for the research design and what can be sacrificed (Benjafield, 1994)? In order to compare exposure outcome results for PTSD, detailed symptom changes for all core symptoms and associated symptoms must be proved. However, by employing too many outcome measures (Marks et al., 1998), as a possible way of assessing more symptoms, not only increases Type I errors, but risks "assessment fatigue" (Harvey et al., 2003).

---

[1] Increased power relates to an increased probability of rejecting the null hypothesis; which is exactly the goal of comparison studies: to find a difference between treatment groups. However, there is a trade-off: increased power relates to an increased probability of making Type I errors (Graziano & Raulin, 2000). The practical value hereof is that extremely high power will suggest differences between treatments that might not be there (Benjafield, 1994).

Since Foa and Meadows (1997) did not suggest a limit to the number of outcome measures that may be used in research, it indicates a limitation in parameter 2 of the Gold Standards. To counter the possible increase in Type I errors by investigating a multitude of symptoms, the number of outcome measures can be reduced and replaced with a multimodal assessment approach. This approach will also help to reduce "assessment fatigue" (Johnson, cited in Harvey et al., 2003).

4.1.2 Estimates of clinical improvement

From chapter 3 it is evident that the identified studies mostly investigated improvement by quantifying symptom change (either as good end-state functioning and/or symptom worsening), diagnosis remission and effect size estimates. Self-rated estimates of improved social functioning were found in only two studies (Marks et al., 1998; Tarrier et al., 1999). Barlow and Hofmann (1997), Harvey et al. (2003), and Graziano and Raulin (2000) considered it important that empirical data be analysed in terms of the practical value of the treatment condition, thereby increasing the external validity of the treatment (King, 1998). Barlow and Hofmann (1997) placed particular emphasis on how participant acceptability of cost, pain, duration of treatment, and side effects influence the feasibility of a treatment, and stressed the need that these variables be assessed.

The Gold Standards do not provide guidelines to estimate significant clinical change. Because the criteria for clinical change are not uniform across treatment outcome studies, comparison of the results of different studies becomes difficult. Over time, treatment outcome research have estimated such change through (a) effect sizes, (b) PTSD diagnosis remission, (c) good end-state functioning, (d) subjective participant ratings of social functioning (both pre- and post-treatment), and (e) significant symptom worsening. It is therefore necessary that the Gold Standards incorporate all of these variables into a uniform criterion to estimate clinical change.

4.1.3 Estimates of treatment adherence

Without adequate estimates of treatment adherence[2], the use of manualised treatments is futile (Foa & Meadows, 1997), as a lack of control over confounding external variables[3] restrict the internal validity

---

[2] The Gold Standards included "treatment integrity" and "treatment adherence" definitions under the auspice of "treatment adherence" (parameter 7). This assignment adopts this viewpoint in its discussions.

[3] In the case of treatment adherence, these external variables include elements of therapist competence.

of the research design (Graziano & Raulin, 2000). One study (Resick et al., 2002) addressed treatment adherence ratings better than the other four studies. The research design utilised independent assessors who rated videotaped sessions according to elements necessary for each therapeutic condition to be completed successfully. The total amount of homework completed was also rated. Limitations of the five outcome studies included (a) the lack of homework adherence ratings (Foa et al., 1999; Tarrier et al., 1999; Taylor et al., 2003) (b) small samples of sessions rated (8% to 9% of sessions) (all studies except Taylor et al., 2003), and (c) incomplete criteria for ratings (not indicating specific treatment components that are checked for) (Tarrier et al., 1999).

Although two studies attempted to rate homework adherence (Marks et al., 1998; Resick et al., 2002), assessment relied on the honesty of participants as to whether or not they actually completed homework assignments. In 1994, Lichstein, Riedel, and Grieve (cited in Harvey et al., 2003) outlined three levels of adherence to treatment. Apart from the treatment be delivered in pure form (level I) and guarantees that the participant has received the treatment (level II), the third level, "enactment" (p. 516), must ensure that the participant has applied the treatment outside of the session. This clearly points to the role of homework exercises.

Exposure therapy makes extensive use of both imaginal and *in vivo* homework exercises. Control for homework is of cardinal importance to the success of exposure treatment, because between-session habituation of the fear network is deemed as one of the three indicators of successful emotional processing (Foa, Riggs, Massie, & Yarkzower, 1995; Marks, 1987). The fact that none of the five studies closely monitored homework exercises indicates that a core aspect of estimating the efficacy of exposure treatment has been neglected.

The APA suggested that therapist competence (including personal attributes and therapist techniques) be rated in treatment efficacy trials (Barlow & Hofmann, 1997). However, few studies attempted to estimate the influence of this variable in-depth (Resick et al., 2002; Taylor et al., 2003). Treatment guidelines provided by the International Society for Traumatic Stress Studies, too, did not expand on necessary therapist competence elements (Foa et al., 2000). Since the biographical and clinical profile of dropouts are closely scrutinised to find possible explanations for treatment failure (Marks et al., 1998), it is surprising that therapist competence do not feature more strongly in the evaluation of treatment success/failure.

The therapeutic alliance has endured as an important variable for psychotherapy change in various schools of psychology (Martin, Garske, & Davis, 2000). A multitude of personal attributes and thera-

pist techniques[4] have been identified that may positively influence the therapeutic alliance (Ackerman & Hilsenroth, 2003), but exposure outcome studies appear to have neglected this invaluable aspect of treatment success. For example, Taylor et al. (2003) only investigated therapist variables as a non-specific treatment component, rated by a blind assessor. Although a blind assessor ensures greater validity, participant ratings of therapeutic alliance factors are necessary to determine how participants view the role of their therapist and the therapeutic relationship, the quality of which could influence treatment adherence.

Of all Gold Standards criteria, parameter 7 ("treatment adherence") appears to be the most vague. Not only are treatment integrity and treatment adherence grouped under the same criterion, but the scope of the definition of treatment adherence do not require estimates of homework adherence, or the influence of therapist competence on treatment adherence.

4.1.4 Inferential statistics for attrition

Reliable and valid statistical procedures are essential to interpret data reliably, especially to deal with attrition that can lead to erroneous interpretations (Graziano & Raulin, 2000). This is especially applicable when participants drop out due to significant factors (e.g., treatment expectations), which affect treatment efficacy directly. By omitting attrition from data analyses, the interpretation of results will be biased, and the generalisability of conclusions restricted when only the results from completers are considered (Harvey et al., 2003). The reasons why a treatment has failed for some prospective participants are as important as why a treatment was successful for others, in order to ensure better external validity for a specific treatment (Barlow & Hofmann, 1997).

Although the usefulness of the statistical analysis of dropout data points is debated (Resick et al., 2002), sound procedures have been used since 1992 to analyse attrition (Mazumbar, Liu, Houck, & Reynolds, 1999), thereby countering arguments not to incorporate ITT samples in analyses. Most studies have used the "last observation carried forward" method up and till now to deal with dropout data. However, in order to prevent misleading findings, the "last observation carried forward" method can be supplemented with mixed-effect linear regression analysis or random regression (Gibbons et al., cited

---

[4] Therapists' personal attributes include being flexible, honest, respectful, trustworthy, confident, warm, interested, and open. Therapist techniques include exploration, reflection, accurate interpretation, facilitating the expression of affect, and attending to the patient's experience (Ackerman & Hilsenroth, 2003).

in Resick et al., 2002).

Foa and Meadows (1997), and Harvey et al. (2003) admitted that the Gold Standards do not provide adequate guidelines for the statistical analyses of outcome data.

4.1.5 *Ethical implementation of exposure treatment*

The ethical implementation of any treatment trial remains a contentious topic (Frueh, Turner, & Beidel, 1995). Of all treatment options for PTSD, exposure therapy receives the most attention regarding its ethical implications and use (Rothbaum & Schwartz, 2002). However, it is still advocated practice to include a wait-list control group in research designs (Foa et al., 2000), despite international initiatives to protect torture victims by forbidding such research practices (Petersen, cited in Paunovic & Öst, 2001). Wait-list groups are deemed important to control for the spontaneous remission of symptoms. Participants' rights to benevolence are the question at hand, and whether or not including a wait-list group infringes on this right. Since none of the outcome studies investigated attrition from wait-list groups in order to establish whether or not the waiting period deterred or harmed them, it is nearly impossible to answer the question.

The Gold Standards do not provide any guidelines on the ethical implementation of exposure treatment for PTSD.

4.2 Conclusions and Recommendations

From the discussion in section 4.1 several limitations have been identified regarding criteria incorporated into Gold Standards. It is concluded that three of the criteria (parameter 1, 2, and 7) do not provide detailed guidelines to ensure internal and external validity of research designs, and warrants recommendations to clarify these criteria. Due to the confounding definition of treatment adherence in parameter 7, it is recommended that parameter 5 (manualised and replicable treatments) necessitates modification. The following recommendations are proposed to modify parameters 1,2, 5, and 7:

Parameter 1: Clearly defined target symptoms. Target symptoms for outcome studies must extend beyond an Axis I level of diagnoses to include the core symptoms of the syndrome under investigation. For example, for PTSD, pre-treatment to follow-up data must include symptom clusters and symptoms of depression and anger, which is crucial for analyses (Foa et al., 1995; Nishith, Hearst, Mueser, & Foa, 1995).

Parameter 2: Reliable and valid measures. Apart from measures of symptom severity and diagnosis, reliable and valid measures must be included for: (a) treatment expectations, (b) treatment integrity, (c) homework adherence, and (d) therapist competence. If such measures do not exist, determined research efforts must address the limitation. Furthermore, the size of the assessment battery must be minimized to avoid "assessment fatigue" and Type I errors. Limiting the research hypotheses to investigate certain core symptoms, or choosing multidimensional measures, may help serve this purpose.

Parameter 5: Specific, manualised and replicable treatments. Parameter 7 provides a confounding definition of treatment adherence by incorporating treatment integrity and treatment adherence under the same definition. This creates uncertainty regarding which estimates need to be included in data analyses. It is recommended that treatment integrity ratings be separately specified and included under parameter 5, which deals with treatment content.

Parameter 7: Treatment adherence ratings. It is recommended that the definition of treatment adherence be simplified to include estimates of: (a) therapy expectations of participants, (b) homework adherence, and (c) therapist competence.

The discussion in section 4.1 also showed limitations in the scope of the Gold Standards. Several methodological limitations are not controlled for by the Gold Standards. Consequently, relying solely on the Gold Standards to provide methodological soundness to a research design will not ensure sufficient control over variables that may restrict generalisation of results. In order to address these limitations three criteria, in addition to the Gold Standards, are proposed.

Parameter 8: Statistical methods. The Gold Standards must be expanded to include criteria for the statistical analyses of therapeutic change as well as methods for additional analyses, such as subgroup analyses and adjusted analyses (Harvey et al., 2003), as well as analyses of attrition. This will also facilitate the comparison of the results from different studies.

Parameter 9: Ethical practice. The APA (Barlow & Hofmann, 1997) emphasises the ethical implementation of treatment procedures. It is recommended that the Gold Standards include guidelines and criteria for ethical practice, particularly in terms of (a) the necessity of a wait-list control group, (b) acceptability of the treatment options to the participant, and (c) choice of treatment for wait-list participants and dropouts. One possibility, if a wait-list group must be included, is to consider the "Minimal Attention" group approach followed by Resick et al. (2002) as an attempt to ensure that participants do not experience adverse effects during the waiting period.

Parameter 10: External validity. It is suggested that the Gold Standards be expanded to provide clear guidelines on estimates of significant clinical improvement. This may so include variables such as (a) effect sizes, (b) PTSD diagnosis remission, (c) end-state functioning, (d) subjective participant ratings of social functioning (both pre- and post-treatment), and (e) symptom worsening. In order to ascertain the reasons for poor or no symptom improvement and/or attrition, it is also recommended that dropouts and poor clinical responders be contacted for a qualitative assessment of possible factors, which might have contributed to their state of functioning, using Barlow and Hofmann's (1997) guidelines on external validity (see Addendum 1).

4.3 Epilogue

Several limitations of this assignment must be acknowledged. *Firstly*, the fact that this study relies on very stringent inclusion criteria and detailed evaluation criteria could be biased as too exclusive, thereby increasing the possibility of Type II errors in drawing conclusions. There might have been significant methodological limitations that were not detected. *Secondly,* this evaluation of the Gold Standards included only exposure outcome studies of PTSD, restricting the generalisation of findings to other psychiatric disorders.

The Gold Standards have emerged as the most widely used criteria for cognitive-behavioural treatment outcome research. Despite the fact that treatment outcome studies adhered to all the required guidelines, important methodological limitations still occurred in research designs and in the reporting of trial results. The most notable limitation is the poor results obtained from estimates of significant clinical change. Although the Gold Standards cannot control the inherent strength of a particular psychological treatment, it can provide the criteria for researchers to estimate valid changes in clinical improvement. This is, however, not necessarily accomplished by means of the Gold Standards, and researchers are left to their own ingenuity to estimate significant clinical change and consequently create a plethora of approaches for such estimates, thus creating dubious interpretations.

The Gold Standards were initially created for treatment outcome research on PTSD. As more empirical research extends the Gold Standards to treatment outcome studies of other psychiatric disorders, it is expected that these criteria will be refined and significant limitations addressed. Together with existing methodology criteria, such as the CONSORT statement, the design and reporting of treatment outcome trials can be put on a still higher level, thereby contributing towards increasing the external validity of psychological treatments.

REFERENCES

Ackerman, S.J., & Hilsenroth, M.J. (2003). A review of therapist characteristics and techniques positively impacting the therapeutic alliance. *Clinical Psychology Review, 23,* 1-33.

Agency for Healthcare Research and Quality (2000). *Outcomes Research.* Fact Sheet. AHRQ Publication No. 00-P011, March 2000. Agency for Healthcare Research and Quality, Rockville, MD. Downloaded 20 October 2004 from the "World Wide Web: http://www.ahrq.gov/clinic/outfact.htm

American Psychiatric Association. (1980). *Diagnostic and Statistical Manual of Mental Disorders* (third edition). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2002). *Diagnostic and Statistical Manual of Mental Disorders* (fourth edition, text revision). Washington, DC: American Psychiatric Association.

Barlow, D.H., & Hofmann, S.G. (1997). Efficacy and dissemination of psychological treatments. In D.M. Clark & C.G. Fairburn (Eds.), *Science and practice of cognitive-behaviour therapy* (pp. 95-117). Oxford: Oxford University Press.

Benjafield, J.G. (1994). *Thinking critically about research methods.* Needham Heights, MA: Allyn and Bacon.

Borkovec, T.D., & Costello, J. (1993). Efficacy of applied relaxation and cognitive behavioral therapy in the treatment of generalized anxiety disorder. *Journal of Consulting and Clinical Psychology, 61,* 611-619.

Boudewyns, P.A., & Hyer, L. (1990). Physiological response to combat memories and preliminary treatment outcome in Vietnam veteran PTSD patients treated with direct therapeutic exposure. *Behavior Therapy, 21,* 63-87.

Boudewyns, P.A., Hyer, L., Woods, M.G., Harrison, W.R., & McCranie, E. (1990). PTSD among Vietnam veterans: an early look at treatment outcome with direct therapeutic exposure. *Journal of Traumatic Stress, 3,* 359-368.

Brümmer, W. (29 October 2003). Medisynebepaling "'n ramp": sonder middele kan selfmoorde en aanvalle kom – kundige. *Die Beeld,* p. 7. (Newspaper article, archives). Downloaded 10 March

2004 from the "World Wide Web": http://152.111.1.251/argief/berigte/beeld/2003/10/29/B3/07/01.html.

Bryant, R.A., Moulds, M.L., Guthrie, R.M., Dang, S.T., & Dixon, R.D.V. (2003). Imaginal exposure alone and imaginal exposure with cognitive restructuring in treatment of posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 71,* 706-712.

Caballo, V.E. (1998). *International handbook of cognitive and behavioral treatments for psychological disorders*. Oxford: Pergamon.

Carlson, J.G., Chemtob, C.M., Rusnak, K., Hedlund, N.L., & Muraoka, M.Y. (1998). Eye movement desensitisation and reprocessing for combat-related posttraumatic stress disorder. *Journal of Traumatic Stress, 11,* 3-24.

Cooper, N.A., & Clum, G.A. (1989). Imaginal flooding as a supplementary treatment for PTSD in combat veterans: a controlled study. *Behavior Therapy, 20,* 381-391.

Echeburúa, E., De Corral, P., Zubizarreta, I., & Sarasua, B. (1997). Psychological treatment of chronic posttraumatic stress disorder in victims of sexual aggression. *Behavior Modification, 21,* 433-456.

Emmelkamp, P.M.G., Visser, S., & Hoekstra, R.J. (1988). Cognitive therapy vs. exposure *in vivo* in the treatment of obsessive-compulsives. *Cognitive Therapy and Research, 12,* 491-495.

Fairbank, J.A., & Keane, T.M. (1982). Flooding for combat-related stress disorders: assessment of anxiety reduction across traumatic memories. *Behavior Therapy, 13,* 499-510.

Foa, E.B., Dancu, C.V., Hembree, E.A., Jaycox, L.H., Meadows, E.A., & Street, G. (1999). A comparison of exposure therapy, stress inoculation training, and their combination for reducing posttraumatic stress disorder in female assault victims. *Journal of Consulting and Clinical Psychology, 67,* 194-200.

Foa, E.B., Keane, T.M., & Friedman, M.J. (2000). Guidelines for treatment of PTSD. *Journal of Traumatic Stress, 13,* 539-563.

Foa, E. B., & Kozak, M.J. (1986). Emotional processing of fear: exposure to corrective information. *Psychological Bulletin, 99,* 20-35.

Foa, E.B., & Meadows, E.A. (1997). Psychosocial treatments for posttraumatic stress disorder: a critical review. *Annual Review of Psychology, 48,* 449-480.

Foa, E.B., Riggs, D.S., Massie, E.D., & Yarkzower, M. (1995). The impact of fear activation and anger on the efficacy of exposure treatment for post-traumatic stress disorder. *Behavior Therapy, 26,* 487-499.

Foa, E.B., Rothbaum, B.O., Riggs, D.S., & Murdock, T.B. (1991). Treatment of posttraumatic stress disorder in rape victims: a comparison between cognitive-behavioral procedures and counselling. *Journal of Consulting and Clinical Psychology, 59,* 715-723.

Foa, E.B., Rothbaum, B.O., & Steketee, G.S. (1993). Treatment of rape victims. *Journal of Interpersonal Violence, 8,* 256-276.

Foa, E.B., Zoellner, L.A., Feeny, N.C., Hembree, E.A., & Alvarez-Conrad, J. (2002). Does imaginal exposure exacerbate PTSD symptoms? *Journal of Consulting and Clinical Psychology, 70, 1022-1028.*

Frank, E., Anderson, B., Stewart, B.D., Dancu, C., Hughes, C., & West, D. (1988). Efficacy of cognitive behavior therapy and systematic desensitisation in the treatment of rape trauma. *Behavior Therapy, 19,* 403-420.

Frueh, B.C., Turner, S.M., & Beidel, D.C. (1995). Exposure therapy for combat-related PTSD: a critical review. *Clinical Psychology Review, 15,* 799-817.

Graziano, A.M.; & Raulin, M.L. (2000). *Research methods: a process of inquiry (fourth edition).* Boston: Allyn and Bacon.

Harvey, A.G., Bryant, R.A., & Tarrier, N. (2003). Cognitive behavior therapy for posttraumatic stress disorder. *Clinical Psychology Review, 23,* 501-522.

Hickling, E.J., & Blanchard, E.B. (1997). The private practice psychologist and manual-based treatments: post-traumatic stress disorder secondary to motor vehicle accidents. *Behaviour Research and Therapy, 35,* 191-203.

Keane, T. M., Fairbank, J. A., Caddell, J. M., & Zimering, R. T. (1989). Implosive (flooding) therapy reduces symptoms of PTSD in Vietnam combat veterans. *Behavior Therapy, 20,* 245-260.

Keane, T.M., & Kaloupek, D.G. (1982). Imaginal flooding in the treatment of a post-traumatic stress disorder. *Journal of Consulting and Clinical Psychology, 50,* 138-140.

King, R. (1998). Evidence-based practice: where is the evidence? The case of cognitive behaviour therapy and depression. *Australian Psychologist, 33,* 83-88.

Lovell, K., Marks, I.M., Noshirvani, H., Thrasher, S., & Livanou, M. (2001). Do cognitive and exposure treatments improve various PTSD symptoms differently? A randomized controlled trial. *Behavioural and Cognitive Psychotherapy, 29,* 107-112.

Marks, I. (1987). Principles of psychological treatment. In I. Marks (ed.), *Fears, phobias and rituals.* New York: Oxford University Press.

Marks, I., Lovell, K., Noshirvani, H., Livanou, M., & Thrasher, S. (1998). Treatment of posttraumatic stress disorder by exposure and/or cognitive restructuring. *Archives of General Psychiatry, 55,* 317-325.

Martin, D.J., Garske, J.P., & Davis, M.K. (2000). Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. *Journal of Consulting and Clinical Psychology, 68,* 438-450.

Maxfield, L., & Hyer, L. (2002). The relationship between efficacy and methodology in studies investigating EMDR treatment of PTSD. *Journal of Clinical Psychology, 58,* 23-41.

Mazumbar, S., Liu, K., Houck, P., & Reynolds, C. (1999). Intent-to-treat analysis for longitudinal clinical trials: coping with the challenge of missing values. *Journal of Psychiatric Research, 33,* 87-95.

Nishith, P., Hearst, D.E., Mueser, K.T., & Foa, E.B. (1995). PTSD and major depression: methodological and treatment considerations in a single case design. *Behavior Therapy, 26,* 319-335.

Paunovic, N., & Öst, L.G. (2001). Cognitive-behavior therapy vs. exposure therapy in the treatment of PTSD in refugees. *Behaviour Research and Therapy, 39,* 1183-1197.

Pitman, R.K., Orr, S.P., Altman, B., & Longpre, R.E. (1999). Emotional processing and outcome of imaginal flooding therapy in Vietnam veterans with chronic posttraumatic stress disorder. *Comprehensive Psychiatry, 37,* 409-418.

Resick, P.A., Nishith, P., Weaver, T.L., Astin, M.C., & Feuer, C.A. (2002). A comparison of cognitive-processing therapy with prolonged exposure and a waiting condition for the treatment of chronic posttraumatic stress disorder in female rape victims. *Journal of Consulting and Clinical Psychology, 70,* 867-879.

Resick, P.A., & Schnicke, M.K. (1992). Cognitive processing therapy for sexual assault victims. *Journal of Consulting and Clinical Psychology, 60,* 748-756.

Richards, D.A., Lovell, K., & Marks, I.M. (1994). Post-traumatic stress disorder: evaluation of a behavioral treatment program. *Journal of Traumatic Stress, 7,* 669-680.

Rothbaum, B. O., & Schwartz, A.C. (2002). Exposure therapy for posttraumatic stress disorder. *American Journal of Psychotherapy, 56,* 59-75.

Salkovskis, P., Atha, C., & Storer, D. (1990). Cognitive-behavioural problem solving in the treatment of patients who repeatedly attempt suicide: a controlled trial. *British Journal of Psychiatry, 157,* 871-876.

Salkovskis, P., Clark, D., & Hackmann, A. (1991). Treatment of panic attacks using cognitive therapy without exposure or breathing retraining. *Behaviour Research and Therapy, 29,* 161-166.

Shapiro, F. (1989). Eye movement desensitisation: a new treatment for post-traumatic stress disorder. Journal of Behavior Therapy and Experimental Psychiatry, 20, 211-217.

Sherman, J.J. (1998). Effects of psychotherapeutic treatments for PTSD: a meta-analysis of controlled clinical trials. *Journal of Traumatic Stress, 11,* 413-435.

Tarrier, N., Pilgrim, H., Sommerfield, C., Faragher, B., Reynolds, M., Graham, E., & Barrowclough, C. (1999). A randomized trial of cognitive therapy and imaginal exposure in the treatment of chronic posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 67,* 13-18.

Taylor, S., Thordarson, D.S., Maxfield, L., Fedoroff, I.C., Lovell, K., & Ogrodniczuk, J. (2003). Comparative efficacy, speed, and adverse effects of three PTSD treatments: exposure therapy, EMDR, and relaxation training. *Journal of Consulting and Clinical Psychology, 71,* 330-338.

Thompson, J.A., Charlton, P.F.C., Kerry, R., Lee, K., & Turner, M. (1995). An open trial of exposure therapy based on deconditioning for post-traumatic stress disorder. *British Journal of Clinical Psychology, 34,* 407-416.

Turner, S.M., Beidel, D.C., & Cooley-Quille, M.R. (1995). Two-year follow-up of social phobics treated with Social Effectiveness Therapy. *Behaviour Research and Therapy, 33,* 553-555.

Wilson, J.P., Friedman, M.J., & Lindy, J.D. (2001). *Treating psychological trauma and PTSD.* New York: The Guilford Press.

ADDENDUM 1

Template for Constructing and Evaluating Psychological Interventions (Adapted from Barlow & Hofmann, 1997)

Internal validity (efficacy)

1. Randomised controlled trials (RANDOMISED CONTROL TRIALS) and randomised comparison trials (Randomised comparison trials)
2. Better than non-specific therapy
3. Better than no therapy
4. Quantified clinical observations
5. Strongly positive clinical consensus
6. Mixed clinical consensus
7. Strongly negative clinical consensus
8. Contradictory evidence

External validity (clinical utility)

1. Feasibility
   - Patient acceptability (cost, pain, duration, side-effects, etc.)
   - Patient-choice in face of relatively equal efficacy
   - Probability of compliance
   - Ease of disseminability (e.g., number of practitioners with competence, requirements for training, opportunities for training, need for costly technologies, or additional support personnel, etc.)
2. Generalisation
   - Patient characteristics
     (i) cultural background issues
     (ii) gender issues
     (iii) developmental issues
     (iv) other relevant patient characteristics
   - Therapist characteristics
   - Issues of robustness when applied in practice settings with different time frames
   - Contextual factors regarding setting in which treatment is delivered
3. Costs and benefits
   - Costs of delivering intervention to individual and society
   - Costs of withholding intervention to individual and society of effective intervention

ADDENDUM 2

Checklist of Items included in the CONSORT Statement (adapted from Harvey et al., 2003)

Section and topic	Description
Title and abstract	Participant allocation to intervention (e.g., "randomised")
Introduction	
Background	Scientific background and explanation of rationale
Methods	
Participants	Inclusion criteria and settings and locations where data were collected
Interventions	Details of interventions intended and method of administration
Objectives	Specific objectives and hypotheses
Outcomes	Defined primary and secondary outcome measures
Sample size	Determination of sample size and explanation of interim analyses and stopping rules
Randomisation sequence generation	Method used to generate random allocation sequence and details of restrictions (e.g., blocking)
Allocation concealment	Method used to implement the random allocation sequence (e.g., central telephone)
Implementation	Who generated allocation, assigned participants?
Blinding (masking)	Did blinding occur and how was the success of blinding evaluated
Statistical methods	Which methods used and why
Results	
Participant flow	Flow of participants through each stage, numbers assigned to each group, treatment completers and analysed for the primary outcome, and deviations from study plan
Recruitment	Dates defining the periods of recruitment and follow-up
Baseline data	Baseline demographic and clinical characteristics of each group
Numbers analysed	Number of participants in each analysis group & whether or not it is "intent-to-treat" analysis

Section and topic	Description
Outcomes and estimation	A summary of results for each group, effect size estimation and its precision (e.g. 95%)
Ancillary analyses	Address multiplicity: report other analysis performed
Adverse events	Report all adverse or side effects
Discussion	
Interpretation	Interpret results, taking hypotheses, bias, and multiplicity into account
Generalisation	External validity of findings
Overall evidence	General interpretation of the results in the context of current evidence