Journal of Affective Disorders 189 (2016) 150–158

Contents lists available at ScienceDirect

Journal of Affective Disorders

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

Research report

Clinical effectiveness of cognitive behavioral therapy for depression in routine care: A propensity score based comparison between randomized controlled trials and clinical practice

Wolfgang Lutz [a,*], Anne-Katharina Schiefele [a,1], Felix Wucherpfennig [a,2], Julian Rubel [a,2], Niklaus Stulz [b,3]

[a] Clinical Psychology and Psychotherapy, Department of Psychology, University of Trier, D-54296 Trier, Germany
[b] Department of Psychology, University of Berne, Fabrikstrasse 8, CH-3012 Berne, Switzerland

ARTICLE INFO

Article history:
Received 3 June 2015
Received in revised form
31 July 2015
Accepted 29 August 2015
Available online 25 September 2015

Keywords:
Randomized controlled trial
Clinical practice
Cognitive behavioral therapy
Depressive disorders
Propensity score matching
Treatment effects

ABSTRACT

Background: The efficacy of cognitive behavioral therapy (CBT) for the treatment of depressive disorders has been demonstrated in many randomized controlled trials (RCTs). This study investigated whether for CBT similar effects can be expected under routine care conditions when the patients are comparable to those examined in RCTs.
Method: $N=574$ CBT patients from an outpatient clinic were stepwise matched to the patients undergoing CBT in the National Institute of Mental Health Treatment of Depression Collaborative Research Program (TDCRP). First, the exclusion criteria of the RCT were applied to the naturalistic sample of the outpatient clinic. Second, propensity score matching (PSM) was used to adjust the remaining naturalistic sample on the basis of baseline covariate distributions. Matched samples were then compared regarding treatment effects using effect sizes, average treatment effect on the treated (ATT) and recovery rates.
Results: CBT in the adjusted naturalistic subsample was as effective as in the RCT. However, treatments lasted significantly longer under routine care conditions.
Limitations: The samples included only a limited amount of common predictor variables and stemmed from different countries. There might be additional covariates, which could potentially further improve the matching between the samples.
Conclusions: CBT for depression in clinical practice might be equally effective as manual-based treatments in RCTs when they are applied to comparable patients. The fact that similar effects under routine conditions were reached with more sessions, however, points to the potential to optimize treatments in clinical practice with respect to their efficiency.

1. Introduction

With a lifetime prevalence of 9.5% depressive disorders are the second most common mental disorder after anxiety disorders (18.1%; Kessler et al., 2005). According to the World Health Organization (WHO) depression is even the leading disorder concerning the overall burden of diseases and it might be the second-leading cause of disability worldwide by 2020 (Murray and Lopez, 1996). Not surprisingly, depression therefore is one of the most intensively studied mental disorders (e.g. Cuijpers et al., 2008, 2014). Actually, more than 350 randomized controlled trails (RCT) on the efficacy of depression treatment have been published. The effects of well-standardized depression treatments found in highly controlled RCTs have to be compared to the effects of depression treatment when delivered under routine care conditions, however. There are several peculiarities of RCTs which aim to strengthen the internal validity of study findings but which may hamper the external validity, that is, transfer of the study's findings to clinical practice:

RCTs usually use highly structured treatment manuals for psychosocial interventions and therapists are intensively trained to ensure that all patients receive a comparable treatment. Therapists in clinical practice may often not follow treatment manuals that strictly. Strict standardization of psychotherapeutic procedures and their one-to-one transfer from RCTs to clinical

* Corresponding author. Fax: +49 651 201 2886.
E-mail addresses: wolfgang.lutz@uni-trier.de (W. Lutz), schiefele@uni-trier.de (A.-K. Schiefele), wucherpfennig@uni-trier.de (F. Wucherpfennig), rubel@uni-trier.de (J. Rubel), niklaus.stulz@psy.unibe.ch (N. Stulz).
[1] Fax: +49 651 201 2886.
[2] Fax: +49 651 201 2886.
[3] Fax: +41 31 631 4155.

http://dx.doi.org/10.1016/j.jad.2015.08.072
0165-0327/

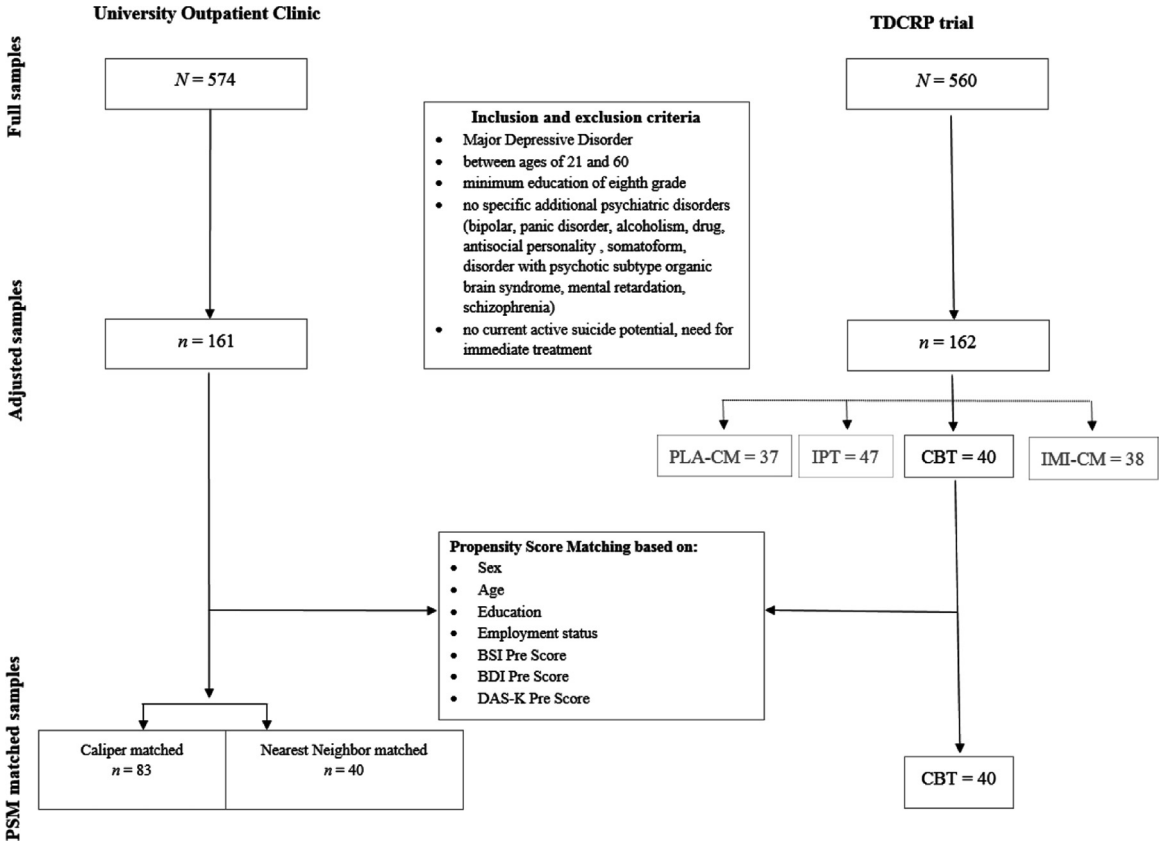

Fig. 1. Flow chart of the full, adjusted and propensity score matched (PSM) samples of the University Outpatient Clinic and the TDCRP trial. Note. TDCRP = Treatment of Depression Collaborative Research Program; CBT = Cognitive Behavioral Therapy; PLA-CM = Placebo plus clinical management; IMI-CM = Imipramine plus clinical management; BSI = Brief Symptom Inventory; BDI = Beck Depression Inventory; DAS-K = Dysfunctional Attitude Scale.

practice is therefore much more difficult in psychotherapy research than for other medical interventions (e.g. pharmacotherapy). Moreover, RCTs usually only include patients who meet a series of highly specific inclusion criteria in order to generate homogenous samples and hence to strengthen the validity of the causal inferences. Combined with the restriction on voluntary patients who accept to be randomly assigned to a treatment condition, these inclusion/exclusion criteria may lead to highly selective samples in RCTs that omit many patients encountered in clinical practice. For instance, studies on antidepressant medications often exclude more than 80% of the patients with a major depression disorder (MDD) due to any non-conformity with the inclusion criteria (e.g. Keitner et al., 2003; Zetin and Hoepner, 2007). While comorbid disorders commonly represent an exclusion criterion in RCTs, patients with more than one mental disorder are frequently seen in clinical practice. Consequently, well-conducted efficacy studies increasingly became criticized in terms of their external validity (Rothwell, 2005), and several efforts have been made to improve the external validity in RCTs. The STAR*D research program, for example, used an equipoised stratified randomized design and gave each patient the possibility to accept the assignment to a particular treatment strategy (e.g., pharmacotherapy and CBT) or decline it and to move to another study arm. This procedure was intended to be more close to what happens in routine care and to reduce the number of non-consenters, resulting in a higher external validity of the study's findings (Warden et al., 2007).

To date, it is generally accepted that both, efficacy (strictly controlled RCTs) and effectiveness studies (studies in naturalistic clinical settings that strengthen external validity at the cost of internal validity) are necessary to evaluate the usefulness of a treatment protocol (Castonguay et al., 2013; Finger and Rand, 2003; Green and Glasgow, 2006; Rothwell, 2005; Taylor and Asmundson, 2008). Results on the transferability of findings from RCTs to naturalistic studies are mixed: while some studies found similar effects (Merrill et al., 2003; Minami et al., 2008), others report that efficacy studies tend to find larger effect sizes than naturalistic studies (Gibbons et al., 2010; Hansen et al., 2002; Weisz et al., 1992). Furthermore, the outcome variance in naturalistic samples tends to be larger than in RCTs (e.g. McEvoy and Nathan, 2007). These findings point to the need for a further investigation of the comparability between treatment effects in RCTs and in naturalistic settings.

We therefore aimed to compare the effects of CBT for patients with MDD in (a) a high-quality RCT (Elkin et al., 1989) and (b) a naturalistic study performed under routine care conditions. As in previous research (e.g. Shadish et al., 1997, 2000; Schindler et al., 2011), we first applied the inclusion/exclusion criteria of the RCT to the sample from routine care to enhance the comparability of the patients examined in both study designs. In addition, we subsequently implemented *propensity score matching* (PSM) to adjust for confounding baseline variables between samples and to match the variable distributions (e.g. Rosenbaum and Rubin, 1983; West et al., 2015).

2. Methods

The current study was based on data from the National Institute of Mental Health Treatment of Depression Collaborative Research Program (TDCRP; Elkin et al., 1989), which was a large multicenter RCT in the US, as well as on naturalistic outcome data,