Research Article   **Open Access**

Brief Cognitive Behavioural Therapy Compared to Optimised General Practitioners' Care for Depression: A Randomised Trial

Schene AH [1,6,7,*], Baas KD[1], Koeter MWJ [1], Lucassen P[3], Bockting CLH[4], Wittkampf KA[1,2], Huyser J[5] and van Weert HC[2]

[1]Department of Psychiatry, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

[2]Department of General Practice, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands

[3]Department of Primary and Community Care, Radboud University Nijmegen, Nijmegen, The Netherlands

[4]Department of Clinical Psychology, University of Groningen, Groningen, The Netherlands

[5]Business Unit Curative Care Arkin, Amsterdam, The Netherlands.

[6]Department of Psychiatry, Radboud University Medical Center, Nijmegen, The Netherlands

[7]Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen, Nijmegen, The Netherlands

*Corresponding author: Aart Schene, Department of Psychiatry, Radboud University Medical Center, Nijmegen, The Netherlands, Tel: 0031 24 36 10674; E-mail: aart.schene@radboudumc.nl

Abstract

Background: How to treat Major Depressive Disorder (MDD) in primary care? Studies that compared (brief) Cognitive Behavioural Therapy (CBT) with care as usual by the General Practitioner (GP) found the first to be more effective. However, to make a fair comparison GP care should be optimised and protocolised according to current evidence based guidelines for depression. So far this has not been the case. We studied whether a protocolised 8 session CBT is more effective than optimised and protocolised GP care (GPC).

Methods: 121 patients with MDD, age 18-70 years, from 40 Dutch general practices, were randomised to either brief CBT or GPC. Assessments were at baseline (t0), 12 weeks (t1) and 52 weeks (t2).

Main outcomes: decrease in depressive symptoms, response and remission on the Hamilton Depression Rating Scale-17 (HDRS-17) and the Patient Health Questionnaire-9 (PHQ-9). (Trial registration: ISRCTN65811640).

Results: Both continuous and dichotomous HDRS-17 and PHQ-9 outcome scores favoured the brief CBT group. Number of treatment contacts and external referrals were not different between groups. GPs prescribed antidepressants (AD) to 48% of GPC patients and to 11% of CBT patients.

Conclusions: Brief CBT by psychologists seems more effective than optimized GPC. Effect sizes comparable to the (statistical significant) results from meta-analyses, together with lower AD prescriptions, are both in favour of brief CBT which might make it a first choice treatment for patients with MDD in general practice.

Keywords: Depressive disorder; Depression; Cognitive behaviour therapy; Antidepressant agents; General practitioner; Randomised controlled trial

Introduction

Major Depressive Disorder (MDD) is a heterogeneous illness with quite different levels of severity and duration, varying numbers of recurrences over the life-time and a ten percent risk for a chronic course. The one year prevalence rates range from 4% to 8% in the general population [1-3] and raise to 12% to 25% among primary care (PC) populations across the world [4]. Related to this heterogeneity, treatment of MDD differs widely in terms of method, intensity, duration, setting and professional(s). Patients with severe, enduring or highly recurrent MDD and those with complicating psychiatric and/or somatic co-morbidity are mostly treated in specialized mental health care settings. However the majority of patients suffer from mild to moderate MDD and those are primarily treated in PC by their General Practitioner (GP) or by mental health professionals, like psychologists [5,6].

For mild to moderate severe MDD treatment guidelines from the National Institute for Health and Clinical Excellence [7] and the American Psychiatric Association [8] recommend low intensity psychological therapy and only in some cases antidepressants (AD; e.g. past history of severe MDD or inadequate response to initial interventions). For moderate to severe MDD these guidelines recommend AD, psychotherapy, or a combination of both. Of the psychotherapeutic interventions for MDD Cognitive Behavioural Therapy (CBT) so far has the most robust evidence [9-11].

CBT for primary care settings is mostly delivered in brief forms, which include six to eight sessions [12-14], about ten sessions less than the 'classic' CBT. In this form CBT is more suitable for PC settings and more acceptable for PC patients. From an explanatory trial perspective brief psychological interventions are better for a comparison with non-psychological treatments than these classical longer ones, where differences in the total amount of treatment and attention may be

problematic when differences in outcome in favour of these psychological treatments have to be interpreted [15].

Effectiveness studies and a recent meta-analysis showed that for MDD patients brief CBT is more effective than treatment as usual by GPs, especially during the first four months [16-20]. A limitation of most comparisons of treatment as usual with an experimental treatment is that the first are not standardised according to recent clinical guideline recommendations with regard to what 'optimal' usual care should include. This unfair competiton may result in a biased outcome; an overestimation of the effect of the experimental treatment (CBT).

Protocolising treatment as usual and improving adherence to this protocol are possible solutions to this methodological issue [21-24]. In this study we did so by standardising the Dutch GP Practice Guideline for MDD [25] and by giving GPs a short training in order to optimise adherence to this protocol. We will refer to this intervention as optimised general practitioners' care (GPC). We hypothesised that brief CBT is more effective than GPC for primary care patients with MDD.

Methods

Setting

The study was conducted in patients of 40 general practices in two parts of the Netherlands; West (Amsterdam region) and East (Nijmegen region). Inclusion took place from January 2007 through August 2010 and follow up assessment till August 2011. The study protocol was approved by the institutional ethics review committees of the Academic Medical Center in Amsterdam and the Radboud University Medical Center in Nijmegen. More details of the study design, including the content of the two treatments, can be found in Baas et al. [26].

Participants and procedure

GPs referred patients with a possible MDD to the study centre. Patients who gave informed consent for the diagnostic phase of the study were assessed by telephone on demographics and clinical characteristics within three days with the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I). Inclusion criteria were: MDD and age between 18 and 70 years. Exclusion criteria were: mental retardation, schizophrenia, bipolar disorder, severe suicidal thoughts, terminal somatic illness, receiving active MDD treatment (a. use of anti-depressive medication, b. psychotherapy, or c. supportive consultations by the GP or social worker) and insufficient comprehension of the Dutch or English language. Eligible patients were next assessed with the Hamilton Depression Rating Scale-17 (HDRS-17) and the Patient Health Questionnaire-9 to establish the severity of MDD. After that they were asked for participation in the trial by their own GP.

Randomisation

After informed consent, the GP performed an internet based randomisation procedure. A centralised computer program generated a block randomisation of four blocks stratified by gender and location (Nijmegen/Amsterdam) where the patient was randomised over the two treatment strategies; optimised general practitioners' care (GPC) or brief cognitive behavioural therapy (CBT). The computer than

generated an outcome on the GP's computer screen and simultaneously send an email of this outcome to the researcher. In this way, patients and investigators were not able to foresee the allocation.

Interventions

General practitioners' care (GPC)

The GPC protocol (developed by HvW and AS; available on request) was based on the Dutch College of General Practitioners Practice Guideline (NHG-Standard) for depressive disorder [25]. GPC had duration of 12 weeks and consisted of monitoring of symptoms, psycho-education, discussion about possible causes or reasons of MDD, problem solving techniques, life style advices and supportive contacts. In accordance with the guidelines GPs were free to decide whether they prescribed an antidepressant (AD) and if so which one. Minimum frequency was one contact (10-15 minutes) every two weeks during the first six weeks and one telephonic contact and one face-to-face evaluation contact during the next 6 weeks. This number could be increased in case of severity of symptoms and/or complaints and/or lack of social support [26].

Brief cognitive behavioural therapy (brief CBT)

The brief CBT protocol (developed by CLHB; available on request) consisted of eight sessions (50 minutes) within the 12 week period. Treatment included behavioural activation and cognitive interventions including identification and challenging of negative thoughts and underlying attitudes and schemas [26]. Therapists were skilled psychologists connected to one of the participating general practices. They all were specifically trained in this brief CBT, had a Master in Clinical Psychology and a four year post academic education in behavioural therapy. They were all members of the Association of Behavioural and Cognitive Therapy. CBT sessions were delivered at the GPs practices.

All brief CBT sessions were audio-taped and treatment integrity was assessed by checking whether the essential ingredients of the intervention were present (behavioural activation, i.e. identification and expanding of potentially pleasant activities and identifying and challenging negative thoughts/ formulating rational thoughts) in a random sample of 10% of each therapists tapes. Therapists had intervision sessions (an organized meeting between colleagues in which performed sessions and related problems are discussed) and supervision sessions (an organized meeting between the therapist and a supervisor, a fully trained cognitive behavioural therapist with a Master in Clinical Psychology and a four year post academic education in Behavioural Therapy and Clinical Psychology). During the 12-week treatment period therapists were asked to refrain from referring the patient to the GP for medication.

Measurements

Patients were assessed at three occasions: before randomisation (baseline; t0); at the end of the treatment period (12 weeks; t1) and at one year follow-up (52 weeks; t2).

Eligibility assessment:

The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) is a semi-structured interview for diagnosing mental disorders according to the DSM-IV criteria [27,28]. Throughout the

study all SCID-I interviewers participated in ongoing training sessions and monthly consensus meetings supervised by an expert psychiatrist (JH) to maximize accuracy and consistency in the administration.

Outcome measures

As primary outcome we used a clinician-rated scale, the Hamilton Depression Rating Scale-17 and a patient-rated scale, the Patient Health Questionnaire-9. For both scales we measured continuous outcomes and dichotomous outcomes (response, remission).

The Hamilton Depression Rating Scale-17 (HDRS-17; [29]) is a well-known clinician rated, semi-structured clinical interview for assessing the severity of depressive disorder which has good psychometric properties [30]. In this study the HDRS was administered by telephone [31] by three independent interviewers, trained by an expert psychiatrist (JH). To maximize accuracy and consistency in the administration we used the same training and supervision procedure as with the SCID-I. The HDRS-17 total score ranges from 0 to 52 points and is categorised as: 8-13, mild; 14-18, moderate; 19-52, severe [32].

The Patient Health Questionnaire-9 (PHQ-9; [33,34]) is the 9 item patient rated subscale of the Primary Care Evaluation of Mental Disorders (PRIME-MD) which has good psychometric characteristics [35] and connects well to primary care [36,37]. In this study the PHQ-9 was send by mail. It evaluates the presence of the nine DSM-IV criteria for a depressive episode. The sum score ranges from 0 to 27 and can be categorised as: 5-9, mild; 10-14, moderate; 15-19, moderately severe; 20-27, severe). The PHQ-9 also offers a diagnostic algorithm. A positive outcome on this algorithm requires that five or more of the nine depressive symptom criteria of the DSM-IV are present more than half the days in the past two weeks (suicidal thoughts count if present at all), and at least one of these five or more symptoms has to be depressed mood or anhedonia.

Dichotomous outcomes

Response: was defined as a total score reduction of >50% on the HDRS-17 or the PHQ-9 [38-40]. Response was assessed only at the end of treatment (12 weeks).

Remission: Using the HDRS-17, remission was defined as a total score of 7 or less. For the PHQ-9 there is no consensus on the definition of remission. To make comparisons with other studies possible we applied two definitions. First, the original validation study of the PHQ-9 [34] recommends that scores of 0-4 are in the minimal depression range and scores of 5-9 are in the mild depression range. We therefore defined remission as a score of <5. Second, the cut-off value more widely used to identify a positive case for depressive disorder is a total score of 10 or higher. We therefore also used a more lenient, remission definition: a PHQ-9 total score <10 [39-41]. Remission was assessed both at 12 weeks (t1) and 52 weeks (t2).

Blinding

Interviewers were kept blind for the group assignment. They only got the telephone number and name of the patient to be interviewed. Before the interview patients were asked "not to reveal their treatment condition". Interviewers were asked in the monthly consensus meetings in how many cases the blinding was broken. They concluded that in 85% to 95% of the cases patients had been able to retain the concealment.

Treatment

In order to optimise GPC each participating GP received training before the start of the study. During this one hour training we educated the GP in the treatment protocol and discussed the content of each of the contacts. After the training we provided a ring binder with the treatment protocol (containing the content of each contact) and informed the GP about the possibility of consulting an independent physician for questions about the treatment protocol during the intervention period.

To monitor the actual content of the provided GPC and potential supplementary treatment next to GPC and brief CBT, GPs received a questionnaire at the end of the 12 week treatment period (t1) in which we asked how many appointments they had had with the patient, whether they prescribed an antidepressant (+ type and dosage) and whether they had referred CBT or GPC patients for supplementary treatment to another health care professional.

Statistical Methods

To evaluate potential differences in baseline characteristics between the two groups we used t-test for independent samples and the chi-square test for categorical variables. When expected cell counts were too small for the chi square test, Fishers exact test was used.

To take potential biased outcomes caused by selective loss to follow-up into account we used multiple imputation (MI) which, assuming missing at random (MAR) for missing values, gives unbiased results with correct standard errors [42]. The results of the multiple imputation analyses (5 imputed datasets) were combined using Rubin's rules [43]. F-test values and degrees of freedom were calculated by the method proposed by Marshall et al. [44]. Since MI based pooled estimates are considered less biased, all analyses are based on the MI data. In the tables we show the estimates based on the imputed data; in the legend of the tables the estimates based on the actual observed data are presented. Furthermore, all analyses were intention to treat. Analyses were carried out with SPSS Statistics 18.0. To guard against an increased family wise error due to multiple testing we corrected the significance threshold from .05 to .01. We used this somewhat more lenient correction than the Bonferoni correction because the 7 tests in our article were not independent (e.g. HDRS score is correlated with PHQ score) which would make a Bonferoni correction to strict and would result in an unnecessary loss of power.

Treatment effect

Continuous outcomes in HDRS-17 and PHQ-9 scores were analyzed with a linear mixed model regression analysis with change between t0 and t1 and change between t0 and t2 as dependent variables, and treatment group, time and the treatment group by time interaction as independent variables, and t0 severity (HDRS total score) as covariate. In this model the main effect of treatment group indicates treatment effect and the treatment by time interaction indicates whether this effect sustains. When treatment by time interaction was significant, we performed planned contrasts to assess whether changes from t0 differed between both groups at t1 and at t2. When the treatment by time interaction was not statistically significant, we fitted a regression model without the interaction term and present the effect estimate based on this model.

Percentage response was analyzed with logistic regression analysis with response in terms of the HDRS-17 or the PHQ-9 as dependent

variable and treatment group as independent variable, and baseline severity as covariate. Percentage remission was analyzed with longitudinal logistic regression analysis, using the SPSS Generalized Estimating Equations programme (GEE) with a logit link function and a binomial error distribution. Dependent variable was remission in terms of the HDRS-17 or the PHQ-9. Independent variables were treatment group, time, the treatment group by time interaction and t0 severity as covariate.

Results

Patient flow and demographic characteristics

Figure 1 shows the patient flow through the study [45]. During the recruitment phase, 175 patients were referred; five patients (2.9%) improved during recruitment and therefore declined participation; 170 could be assessed for eligibility. Of the 170 patients 34 (20%) patients did not meet these criteria, eight (4.7%) patients refused randomization due to treatment preference, six (3.5%) patients said they were improved and therefore declined randomization, and one (0.6%) patient did not accept the diagnosis of MDD; 121 (71.2%) met the inclusion criteria and agreed to participate and to be randomised. Data on outcome were obtained for 94 (78%) patients at 12 weeks and for 77 (64%) patients at 52 weeks.
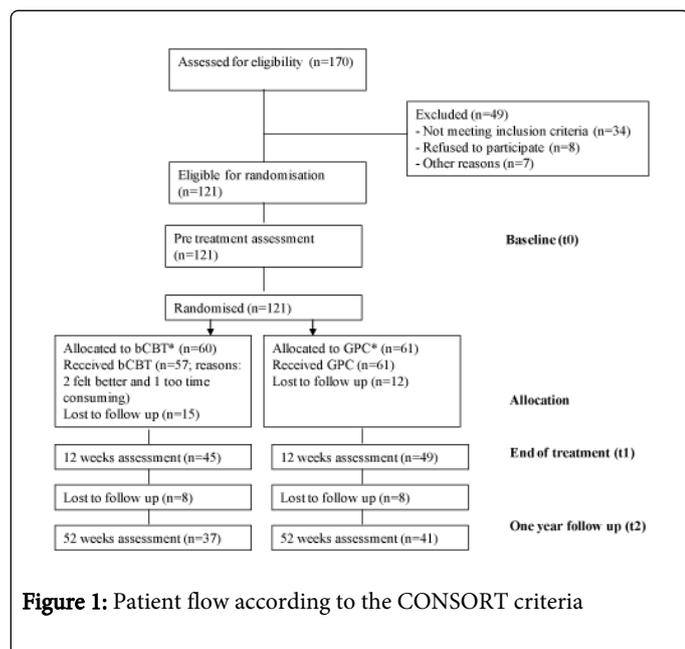

Figure 1: Patient flow according to the CONSORT criteria

Mean age of the total sample (n=121) was 45.1 years (s.d 11.2) and 66.9% of the patients were women (Table 1). Mean HDRS-17 score was 19.1 (s.d 6.1) and mean PHQ-9 score was 17.8 (s.d 4.6), both on the border of moderate and severe MDD. Almost half of the patients (45.7%) had a recurrent MDD and one third (31.4%) suffered from a co-morbid DSM-IV axis-I disorder (substance related disorder, anxiety disorder, somatoform disorder). Both groups were comparable on all demographic and clinical characteristics, both in terms of statistical significance and effect size.

[a]GPC=general practitioners' care; CBT=cognitive behavioural therapy

[b]Educational attainment: three categories according to the Dutch Educational System

[c]Including: substance related disorders, anxiety disorder, and somatoform disorders

	GPC [a] (n=61)	CBT [a] (n=60)
Demographic characteristics	45.3 (11.0)	44.9 (11.6)
Age, mean (s.d)		
Female %	72.1	61.7
Living situation %		
- Alone	25.0	21.4
-With partner (with or without child)	50.1	55.4
-With child (without partner)	18.3	10.7
-With parent(s)	3.3	7.1
-Other	3.3	5.4
Caucasian %	78.3	78.6
Marital status %		
-Not married	45.0	42.8
-Married	36.7	41.1
-Divorced	11.7	14.3
-Widowed	6.6	1.8
Primary role %		
- Employed	56.7	62.5
- Student	0.0	1.8
-Unemployed	26.7	23.2
- Retired	3.3	1.8
- Other	13.3	10.7
Educational attainment[b] %		
- Low	31.7	33.9
- Medium	48.3	41.1
- High	20.0	25.0
Clinical characteristics		
Recurrent depression %	39.6	51.9
Severity depression (HDRS) %		
- Mild (total score 8-13)	42.6	46.7
- Moderate (total score 14-18)	36.1	38.3
- Severe (total score 19-52)	21.3	15.0
HDRS-17, mean (s.d)	19.8 (6.4)	18.3 (5.8)
PHQ-9, mean (s.d)	18.3 (4.0)	17.4 (5.1)

Co-morbid axis-I disorders, current[c] %	32.8	30.0

Table 1: Baseline demographic and clinical characteristics (n=121)

Treatment Effect

Continuous outcomes

Table 2 shows that from t0 to t2 mean HDRS-17 scores declined from 19.8 to 10.1 in the GPC group and from 18.3 to 8.2 in the brief CBT group. Mean PHQ-9 scores declined from 18.3 to 9.4 in the GPC group and from 17.4 to 7.1 in the brief CBT group. No significant treatment by time interaction was found, neither for the HDRS-17, nor for the PHQ-9.

Therefore we restricted ourselves to the regression model with only the main effects. According to this model patients in the brief CBT group improved on average 2.17 points more on the HDRS-17 total score than patients in the GPC group (F(1, 18) = 1.96; 95% CI -0.86-4.37; p=.178; Cohen's d: .36). Patients in the brief CBT group improved 2.38 points more on the PHQ-9 total score compared to patients in the GPC group (F(1, 320) = 6.71; 95% CI 0.59- 4.78; p=.010; Cohen's d: .51).

	GPC[c] (n=61)	CBT[c] (n=60)	F (df1,df2)[d,e]	P[d,e]
HDRS-17(mean, s.d.)			1.96 (1,18)	0.178
Baseline (t0)	19.8 (6.4)	18.3 (5.8)		
12 weeks (t1)	13.0 (8.5)	9.7 (7.4)		
52 weeks (t2)	10.1 (9.0)	8.2 (6.9)		
PHQ-9(mean, s.d.)			6.71 (1,320)	0.010
Baseline (t0)	18.3 (4.0)	17.4 (5.1)		
12 weeks (t1)	11.2 (5.9)	8.3 (5.5)		
52 weeks (t2)	9.4 (5.8)	7.1 (5.2)		

Table 2: Continuous depression outcome[a, b]

[a]Presented are multiple imputation based pooled estimates

[b]Observed mean scores HDRS-17 (t0, t1, t2):

- GPC: 19.8 (n=61), 12.3 (n=49), 9.0 (n=41)

- CBT: 18.3 (n=60), 9.2 (n=45), 8.3 (n=37)

Observed mean scores PHQ-9 (t0, t1, t2):

- GPC: 18.3 (n=60), 11.1 (n=43), 8.9 (n=37)

- CBT: 17.4 (n=58), 7.7 (n=34), 6.7 (n=32)

[c]GPC=general practitioners' care; CBT=cognitive behavioural therapy

[d]Linear mixed model regression analysis with change between baseline (t0) and 12 weeks (t1) and change between baseline (t0) and 52 weeks (t2) as dependent variables; treatment group, time and the treatment group by time interaction as independent variables

[e]Pooled multivariate test according to Marshall et al [44]

Dichotomous outcomes

Response: Table 3 shows that after 12 weeks of treatment on the HDRS-17 a response was shown by 34.1% of the GPC group and by 48.7% of the brief CBT group. For the PHQ-9 responses were respectively 39.3% and 57.0%. According to the logistic regression model CBT and GPC patients did not differ in response percentage on the HDRS-17 (odds ratio 1.90; F (1, 1183)= 2.70; 95% CI 0.88- 4.09; p=.100; Cohen's h: .30), nor on the PHQ-9 (odds ratio 2.22; F (1, 83)=3.42; 95% CI 0.95- 5.14; p=.068; Cohen's h: .36).

Remission: Based on HDRS-17 scores 32.5% of the patients in the GPC group and 51.7% in the brief CBT group could be considered a remitter at 12 weeks and respectively 44.6%and 55.7% at 52 weeks. The logistic regression without treatment by time interaction showed no significant relation between treatment type and HDRS-17 remission (odds ratio 1.97; F(1, 34)= 2.49; 95% CI 0.87- 3.70; p=.124). For the PHQ-9 cut off <10 these figures were respectively 42.6% and 57.3% at 12 weeks and 54.8% and 71.7% at 52 weeks. Logistic regression without treatment by time interaction showed no significant relation between treatment type and PHQ-9 remission when using <10 as cut off score (odds ratio 2.14; F(1, 268)= 3.52; 95% CI 0.95-4.83; p=0.062), nor when remission was defined as a PHQ score <5 (odds ratio 1.87; F (1,67)=3.42; 95% CI 0.97-3.63; p=0.069). In the latter case MI based remission rates were respectively 10.8% (GPC) and 30.0% (CBT) at 12 weeks and 23.0% and 33.0% at 52 weeks.

	GPC[c] (n=61)	CBT[c] (n=60)	F(df1,df2)[d,e]	P[d,e]	effect size[h]
Response[f](at 12weeks; t1)					
HDRS-17	34.1%	48.7%	2.70(1,1183)	0.100	0.30
PHQ-9	39.3%	57.0%	3.42(1,83)	0.068	0.36
Remission[g]					
HDRS-17			2.49(1,34)	0.124	
12 weeks (t1)	32.5%	51.7%			0.39
52 weeks (t2)	44.6%	55.7%			0.22
PHQ-9 (<10)			3.52(1,268)	0.062	
12 weeks (t1)	42.6%	57.3%			0.30
52 weeks (t2)	54.8%	71.7%			0.35
PHQ-9 (<5)			3.42(1,67)	0.069	
12 weeks (t1)	10.8%	30.0%			0.49
52 weeks (t2)	23.0%	33.0%			0.22

Table 3: Dichotomous depression outcome: response and remission[a,b]

[a]Presented are multiple imputation based pooled estimates

[b]Observed percentages response (12 weeks; t1):

- HDRS-17: GPC 36.7%, CBT 51.1%

- PHQ-9: GPC 39.5%, CBT: 61.8

Observed percentages remission (12 weeks; t1):

- HDRS-17: GPC 36.7%, CBT 55.6%

- PHQ-9 (<10): GPC 44.2%, CBT 61.8%

- PHQ-9(<5): GPC 11.6%, CBT 35.3%

Observed percentages remission (52 weeks; t2):

- HDRS-17: GPC 51.2%, CBT 56.8%

- PHQ-9 (<10): GPC 59.5%, CBT 75.0%

- PHQ-9 (<5): GPC 27.3%, CBT 37.5%

[c]GPC=general practitioners' care; CBT=cognitive behavioural therapy

[d]Response: logistic regression analysis with response (Y/N) as outcome variable and treatment as dependent variable; Remission: GEE analysis with remission (Y/N) as dependent and treatment and time as independent variables (treatment by time interaction was not significant)

[e]Pooled multivariate test according to Marshall et al. [44]

[f]Response: > 50% decrease of the baseline score (t0)

[g]Remission: HDRS-17 score ≤7; PHQ-9 score <10; PHQ-9 score<5

[h]Cohen's d (0.20 = small; 0.50 = medium, 0.80 = large)

Adherence and Antidepressant Use

Patients in the brief CBT group received a mean of 6.1 sessions with the psychologist (range 0-8; s.d 2.7). Three patients did not attend any treatment session. The most common reason for termination of therapy was lack of motivation to attend the sessions. Patients in the GPC group received a mean of 4.9 appointments with their GP (range 1-12; s.d 2.2). All patients attended at least one GP-appointment.

Treatment integrity assessment of CBT showed that the essential ingredients of the intervention (i.e. behavioural activation and challenging negative thoughts/ formulating rational thoughts) were present. No interference in the work procedures of the psychologists was necessary.

GPs prescribed antidepressants to 48% of the GPC patients and to 11% of the CBT patients. In total one in seven patients received an outside referral for additional mental health treatment (11% in brief CBT and 18% in GPC).

Discussion

We investigated whether brief cognitive behavioural therapy (CBT) applied by psychologists was more effective than optimised and protocolised general practitioners' care (GPC), based on a GP treatment guideline for primary care patients with MDD. Over the full 52 weeks follow up the depression severity in both groups reduced to about half of the baseline value. Both on the HDRS-17 interview and the PHQ-9 self-rating about 70% of this improvement was between baseline (t0) and the end of treatment at 12 weeks (t1). Our study shows that CBT is at least as effective as GPC (both treatments did not differ on the dichotomous outcomes: response at t1 on HDRS and PHQ-9, and remission on t1 and t2 on HDRS and PHQ-9) and gives an indication that CBT is perhaps more effective (improvement on the mean PHQ-9 score differed significantly between both groups in favour of brief CBT [effect size: .51] and the combined overall trend appears to consistently favour brief CBT). Till now effectiveness

studies and a recent meta-analysis showed that brief CBT is more effective than treatment as usual by GPs. We studied an optimized form of GP care and conclude that brief CBT is just as good and perhaps better.

It cannot be ruled out that due to the sample size our tests had insufficient power to detect effects of the observed magnitude. This is supported by the fact that our effect sizes (most between .30 and .50) are comparable in magnitude to the statistical significant pooled effect sizes of three meta-analyses of CBT versus usual care [20,46,47] which were all in favour of CBT and were in the range of 0.33 to 0.42. Given the effect sizes it is not unreasonable to expect that we might have found more statistical significant effects of treatment type with a larger sample size although such a larger sample size might reduce effect sizes [11].

For the PHQ-9 we used two remission rates and found that those based on the definition of a PHQ-9 cut off <10 were more comparable to the HDRS-17 remission rates than those based on the definition of a PHQ-9 cut off <5. Compared to HDRS-17 the first one (<10) seems to overestimate remission by some 10%, while the second one (<5) underestimates remission by some 20%.

Our remission data at 52 weeks follow up need some explanation. First it is known that 50% of depressive episodes will be in remission at three months, but another 25% will last longer than 12 months [48]. It is also known that during the first year the proportion of patients relapsing/recurring is 25% [49]. To consider a new period of depression as a real recurrence the DSM IV states that between episodes 'there must be an interval of at least 2 consecutive months in which criteria are not met for a Major Depressive Episode'. Unfortunately we were not able to assess relapses or recurrences of depression in that detail over the t1-t2 period. We only know that non-remission (considering PHQ-9 cut off <10 and HDRS-17) at 52 weeks was in the 28-44% range for CBT and in the 45-55% range for GPC. Whether this non-remission was related to 'real' non-remission, relapse or recurrence is not known.

This issue of remission/recurrence of course is interesting because one important advantage of CBT in comparison with AD is its prophylactic effect on depression recurrence after acute CBT/AD treatment has been stopped [49-51]. However in almost all studies on this prophylactic aspect, CBT was of the 16-18 sessions type, as was common for the acute treatment of CBT, and so it remains unclear whether the brief CBT we studied also has an enduring effect.

Patients in de brief CBT group received six therapy sessions (sometimes called ultra-brief CGT; [52]), about two sessions less than the intended and protocolised eight sessions which is in line with other primary care studies on brief CBT in primary care [13,20]. The five (face to face or telephonic) appointments of the GPC patients correspond to the recommended number in the protocol. During the 12 week treatment phase patients in the GPC group received one session less than the brief CBT patients which suggest a comparable number of professional contacts in both groups. Furthermore the referral rate to other mental health treatments was also comparable; one in seven patients in both groups.

Anti-depressants (AD) were one of the treatment options in the GPC. The GPs were educated in the treatment protocol and they received feedback of the SCID-I and HDRS interview results (i.e. diagnosis and severity of MDD). By this procedure they were in an optimal condition to decide to prescribe ADs in accordance with the primary care guidelines. This finally resulted in 48% of AD

prescriptions in the GPC group. We think this percentage is in accordance with the depressions severity at baseline, but rather low compared to other GP studies where 49-96% of the patients received ADs [17,53-55].

Interestingly we have to conclude that the somewhat better depression outcome was accomplished with a small percentage of AD prescriptions (11%) in the brief CBT group. Since studies showed that patients generally prefer psychological therapy to pharmacological therapy [56-59], this low AD prescription may be another advantage of brief CBT. However, we have no assessment of patient' preference or satisfaction, and so this advantage is not proven in our sample.

Earlier studies on brief CBT in depressed PC patients are scarce and showed brief CBT to be more effective than treatment as usual, especially in the short term, but the advantages were small [16-18,47]. Our study differs from these earlier studies in several ways. First, we optimised treatment as usual by using the treatment protocol and the GP training. By this, we think our study design results in a more valid comparison and consequently more valid estimate of the differences in treatment effects between brief CBT and GPC. To our knowledge only one other study also educated GPs; Conradi et al. [19] invited GPs to attend a 2-hour booster session about guidelines for the treatment of depression. Second, the number of offered sessions of the brief CBT differed. In two of the earlier studies brief CBT consisted of ten or more sessions (range 10-12 sessions; [16,19]), while we finally used the six sessions found in clinical practice. Furthermore, the content of the brief CBT differed. Scott et al. [17], for instance, used mainly cognitive techniques instead of a mixture of cognitive and behavioural aspects as we did. Finally, the way the severity of depression was assessed differed. Only one study used both a patient rated and a clinician rated instrument [17]. The other studies used either a clinician rated instrument (HDRS-17; [16]) or a patient rated instrument (Beck Depression Inventory [60]). A recent meta-analysis showed that clinician-rated and patient-reported measures of improvement in depression studies are not equivalent and the authors recommended using both [14] as we did.

Strengths and limitations

Our study has limitations. First, one could argue that a control or treatment as usual group should have been included. We think that CBT had already been proven effective as compared to care as usual [20] and so we considered a placebo or control group not necessary and even unethical. Second, we obtained no information about harms (adverse effects) of the GPC while, for example, side effects of medication can have an influence on whether an intervention will be acceptable. Third, although we tried hard, we were not successful to measure adherence to the GPC protocol because this measure was too time-consuming for busy GPs. However, we monitored the actual content of the provided GPC by sending each GP a short questionnaire at the end of the intervention period in which we asked how many appointments they had had with the patient, whether they prescribed an antidepressant and whether they referred patients to another health care professional. This provided a global estimate of adherence to the protocol. Fourth, we did not measure the total amount of treatment patients in both conditions received between 12 weeks and follow-up at 52 weeks. Finally, we did use multiple imputations (MI) which requires the missing mechanism to be missing at random (MAR). Unfortunately, the MAR assumption is not testable [62]. Serious violations of MAR are, however, unlikely and lesser violations are only problematic in specific situations that are also

rare [63-65]. Alternatives to MI like complete cases analysis or mixed model regression analysis on the observed data either are much more likely to lead to biased results (complete cases analysis) or also require MAR (mixed model regression).Even though we used the best method we cannot completely rule out that the MAR assumption violations induced some bias. However, even in this case the results after MI are most likely less biased than the results from the observed data.

Strengths of this study are the adequate randomisation procedure, the follow-up of 52 weeks, the assessment of AD prescriptions and the detailed analyses of treatment effects. We have tried to improve generalizability by including a great number of general practices from two geographical areas and by using lenient inclusion criteria, permitting co-morbidity [61,62]. Finally the optimisation and standardisation of the GP treatment and the limitations for external referrals allowed a fairer comparison.

Conclusion

We conclude that compared to optimised GPC the brief CBT we developed is probably more successful on depression outcome. This finding, together with a substantial reduction of AD prescriptions in the brief CBT group and the similarity of our effect sizes to the (statistical significant) pooled effect sizes of three meta-analyses of CBT versus usual care, makes brief CBT a good first choice treatment for PC patients with MDD. For the majority of patients a primary care intervention seemed sufficient. One in seven patients was referred for additional therapy. However further trials are needed to support the probable superiority of brief CBT in comparison to optimized GP care and to examine the relative cost-effectiveness.

Funding

Declaration of interest:

None.

Acknowledgments:

References

1. Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, et al. (2004) Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. Acta Psychiatr Scand Suppl : 21-27.

2. Demyttenaere K, Bruffaerts R, Posada-Villa J, Gasquet I, Kovess V, et al. (2004) Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. JAMA 291: 2581-2590.

3. Kessler RC (2007) The global burden of anxiety and mood disorders: putting the European Study of the Epidemiology of Mental Disorders (ESEMeD) findings into perspective. J Clin Psychiatry 68 Suppl 2: 10-19.

4. Mitchell AJ, Rao S, Vaze A (2011) International comparison of clinicians' ability to identify depression in primary care: meta-analysis and meta-regression of predictors. Br J Gen Pract 61: e72-80.

5. Aikens JE, Kroenke K, Nease DE Jr, Klinkman MS, Sen A (2008) Trajectories of improvement for six depression-related outcomes. Gen Hosp Psychiatry 30: 26-31.

6. Bijl RV, Ravelli A (2000) Psychiatric morbidity, service use, and need for care in the general population: results of The Netherlands Mental Health Survey and Incidence Study. Am J Public Health 90: 602-607.

7. NICE (2010) The NICE Guideline on the Treatment and Management of Depression in Adults (updated edn). London, National Institute for Health and Clinical Excellence.

8. APA (2010) Practice Guideline for the Treatment of Patients with Major Depressive Disorder (3rd edn). Washington DC, American Psychiatric Association.

9. Gloaguen V, Cottraux J, Cucherat M, Blackburn IM (1998) A meta-analysis of the effects of cognitive therapy in depressed patients. J Affect Disord 49: 59-72.

10. Cuijpers P, van Straten A, Andersson G, van Oppen P (2008) Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. J Consult Clin Psychol 76: 909-922.

11. Barth J, Munder T, Gerger H, Nüesch E, Trelle S, et al. (2013) Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. PLoS Med 10: e1001454.

12. Churchill R, Hunot V, Corney R, Knapp M, McGuire H, et al. (2001) A systematic review of controlled trials of the effectiveness and cost-effectiveness of brief psychological treatments for depression. Health Technol Assess 5: 1-173.

13. Stiles WB, Barkham M, Connell J, Mellor-Clark J (2008) Responsive regulation of treatment duration in routine practice in United Kingdom primary care settings: replication in a larger sample. J Consult Clin Psychol 76: 298-305.

14. Cuijpers P, Li J, Hofmann SG, Andersson G (2010) Self-reported versus clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: a meta-analysis. Clin Psychol Rev 30: 768-778.

15. Cuijpers P, Huibers M, Ebert DD, Koole SL, Andersson G (2013) How much psychotherapy is needed to treat depression? A metaregression analysis. J Affect Disord 149: 1-13.

16. Scott AI, Freeman CP (1992) Edinburgh primary care depression study: treatment outcome, patient satisfaction, and cost after 16 weeks. BMJ 304: 883-887.

17. Scott C, Tacchi MJ, Jones R, Scott J (1997) Acute and one-year outcome of a randomised controlled trial of brief cognitive therapy for major depressive disorder in primary care. Br J Psychiatry 171: 131-134.

18. Ward E, King M, Lloyd M, Bower P, Sibbald B, et al. (2000) Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy, and usual general practitioner care for patients with depression. I: clinical effectiveness. BMJ 321: 1383-1388.

19. Conradi HJ, de Jonge P, Ormel J (2008) Cognitive-behavioural therapy v. usual care in recurrent depression. Br J Psychiatry 193: 505-506.

20. Cape J, Whittington C, Buszewicz M, Wallace P, Underwood L (2010) Brief psychological therapies for anxiety and depression in primary care: meta-analysis and meta-regression. BMC Med 8: 38.

21. Donoghue JM, Tylee A (1996) The treatment of depression: prescribing patterns of antidepressants in primary care in the UK. Br J Psychiatry 168: 164-168.

22. Roth A, Fonagy P (1996) What works for whom? A critical review of psychotherapy research. New York, Guilford Press.

23. Blenkiron P (1998) Does the management of depression in general practice math current guidelines? Primary Care Psychiatry 4:121-125.

24. Mant D (2008) The problem with usual care. Br J Gen Pract 58: 755-756.

25. Marwijk HWJ van, Grundmeijer HGLM, Bijl D, Gelderen MG van, Haan M de, et al. (2003) NHG- Standaard Depressieve stoornis (eerste herziening). Huisarts en Wetenschap 46:614-633.

26. Baas KD, Koeter MW, van Weert HC, Lucassen P, Bockting CL, et al. (2010) Brief cognitive behavioral therapy compared to general practitioners care for depression in primary care: a randomized trial. Trials 11: 96.

27. Spitzer RL, Williams JB, Gibbon M, First MB (1992) The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. Arch Gen Psychiatry 49: 624-629.

28. First MB, Spitzer RL, Gibbon M, Williams JBW (1999) Structured Clinical Interview for DSM-IV Axis I Disorders- Patient Edition (SCID-I/P version 2.0). Lisse, Swets & Zeitlinger BV.

29. Hamilton M (1960) A rating scale for depression. J Neurol Neurosur PS 23:56-62.

30. Rabkin JG, Klein DF (1987) The clinical measurement of depressive disorders; in Marsella AJ, Hirschfeld RMA, Katz MM (eds): The measurement of depression. New York, Guilford Press, pp 30-83.

31. Simon GE, Revicki D, VonKorff M (1993) Telephone assessment of depression severity. J Psychiatr Res 27: 247-252.

32. Hamilton M (1967) Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 6: 278-296.

33. Spitzer RL, Kroenke K, Williams JB (1999) Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 282: 1737-1744.

34. Kroenke K, Spitzer RL, Williams JB (2001) The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 16: 606-613.

35. Kroenke K, Spitzer RL, Williams JB, Löwe B (2010) The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. Gen Hosp Psychiatry 32: 345-359.

36. Wittkampf KA, Naeije L, Schene AH, Huyser J, van Weert HC (2007) Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. Gen Hosp Psychiatry 29: 388-395.

37. Nease DE Jr, Nutting PA, Dickinson WP, Bonham AJ, Graham DG, et al. (2008) Inducing sustainable improvement in depression care in primary care practices. Jt Comm J Qual Patient Saf 34: 247-255.

38. Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, et al. (1991) Conceptualization and rationale for consensus definitions of terms in major depressive disorder. Remission, recovery, relapse, and recurrence. Arch Gen Psychiat 48:851-855.

39. Löwe B, Schenkel I, Carney-Doebbeling C, Göbel C (2006) Responsiveness of the PHQ-9 to Psychopharmacological Depression Treatment. Psychosomatics 47: 62-67.

40. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K (2004) Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care 42: 1194-1201.

41. Zimmerman M (2012) Symptom severity and guideline-based treatment recommendations for depressed patients: implications of DSM-5's potential recommendation of the PHQ-9 as the measure of choice for depression severity. Psychother Psychosom 81: 329-332.

42. Donders AR, van der Heijden GJ, Stijnen T, Moons KG (2006) Review: a gentle introduction to imputation of missing values. J Clin Epidemiol 59: 1087-1091.

43. Mehrotra DV, Li X, Liu J, Lu K (2012) Analysis of longitudinal clinical trials with missing data using multiple imputation in conjunction with robust regression. Biometrics 68: 1250-1259.

44. Marshall A, Altman DG, Holder RL, Royston P (2009) Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Med Res Methodol 9: 57.

45. Schultz KF, Altman DG, Moher D, the CONSORT Group (2010) CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Medicine 8:18.

46. Bortolotti B, Menchetti M, Bellini F, Montaguti MB, Berardi D (2008) Psychological interventions for major depression in primary care: a meta-analytic review of randomized controlled trials. Gen Hosp Psychiatry 30: 293-302.

47. Cuijpers P, van Straten A, van Schaik A, Andersson G (2009) Psychological treatment of depression in primary care: a meta-analysis. Br J Gen Pract 59: e51-60.

48. Spijker J, de Graaf R, Bijl RV, Beekman AT, Ormel J, et al. (2002) Duration of major depressive episodes in the general population: results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Br J Psychiatry 181: 208-213.

49. Vittengl JR, Clark LA, Dunn TW, Jarrett RB (2007) Reducing relapse and recurrence in unipolar depression: a comparative meta-analysis of cognitive-behavioral therapy's effects. J Consult Clin Psychol 75: 475-488.

50. Hollon SD, Stewart MO, Strunk D (2006) Enduring effects for cognitive behavior therapy in the treatment of depression and anxiety. Annu Rev Psychol 57: 285-315.

51. Cuijpers P, Hollon SD, van Straten A, Bockting C, Berking M, et al. (2013) Does cognitive behaviour therapy have an enduring effect that is superior to keeping patients on continuation pharmacotherapy? A meta-analysis. BMJ Open 3.

52. Rieu J, Bui E, Rouch V, Faure K, Birmes P, et al. (2011) Efficacy of ultrabrief cognitive and behavioural therapy performed by psychiatric residents on depressed inpatients. Psychother Psychosom 80: 374-376.

53. Schulberg HC, Block MR, Madonia MJ, Scott CP, Rodriguez E, et al. (1996) Treating major depression in primary care practice. Eight-month clinical outcomes. Arch Gen Psychiatry 53: 913-919.

54. King M, Sibbald B, Ward E, Bower P, Lloyd M, et al. (2000) Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care. Health Technology Assessment 4:1-83.

55. Dumesnil H, Cortaredona S, Verdoux H, Sebbah R, Paraponaris A, et al. (2012) General practitioners' choices and their determinants when starting treatment for major depression: a cross sectional, randomized case-vignette survey. PLoS One 7: e52429.

56. Priest RG, Vize C, Roberts A, Roberts M, Tylee A (1996) Lay people's attitudes to treatment of depression: results of opinion poll for Defeat Depression Campaign just before its launch. BMJ 313: 858-859.

57. Jorm AF (2000) Mental health literacy. Public knowledge and beliefs about mental disorders. Br J Psychiatry 177: 396-401.

58. van Schaik DJ, Klijn AF, van Hout HP, van Marwijk HW, Beekman AT, et al. (2004) Patients' preferences in the treatment of depressive disorder in primary care. Gen Hosp Psychiatry 26: 184-189.

59. Høifødt RS, Strøm C, Kolstrup N, Eisemann M, Waterloo K (2011) Effectiveness of cognitive behavioural therapy in primary health care: a review. Fam Pract 28: 489-504.

60. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J (1961) An inventory for measuring depression. Arch Gen Psychiat 4:561-571.

61. Fairhurst K, Dowrick C (1996) Problems with recruitment in a randomized controlled trial of counselling in general practice: causes and implications. J Health Serv Res Policy 1: 77-80.

62. Hunt CJ, Shepherd LM, Andrews G (2001) Do doctors know best? Comments on a failed trial. Med J Aust 174: 144-146.

63. Enders GK (2010) Applied missing data analysis. New York, The Guilford press.

64. Graham JW, Hofer SM, Donaldson SI, MacKinnon DP, Schafer JL (1997) Analysis with missing data in prevention research: in: Bryant KJ, Windle M, West G (eds): The science of prevention: methodological advances from alcohol and substance abuse research. Washington DC, American Psychological Association, pp 325-366

65. Schafer JL, Graham JW (2002) Missing data: our view of the state of the art. Psychol Methods 7: 147-177.