Effects of pharmacotherapy and psychotherapy in depressed primary-care patients: a randomized, controlled trial including a patients’ choice arm

Ulrich Hegerl1, Martin Hautzinger2, Roland Mergl1, Ralf Kohnen3, Michael Schütze1, Winfried Scheunemann1, Antje-Kathrin Allgaier5, James Coyne6 and Verena Henkel7

1 Department of Psychiatry, University of Leipzig, Leipzig, Germany
2 Department of Psychology, Eberhard-Karls-University Tuebingen, Tuebingen, Germany
3 IMEREM Institute for Medical Research Management and Biometrics (IMEREM), Nuremberg, Germany
4 20 rue Louvière, L-1946 Luxembourg, Luxembourg
5 Department of Child and Adolescent Psychiatry, Ludwig-Maximilians-University Munich, Munich, Germany
6 Department of Psychiatry, University of Pennsylvania Health System, Philadelphia, PA, USA
7 Department of Psychiatry, Ludwig-Maximilians-University Munich, Munich, Germany

Abstract

Mild depressive syndromes are highly prevalent among primary-care patients. Evidence-based treatment recommendations need to be derived directly from this diagnostically heterogeneous group. The primary aim was to assess the efficacy of sertraline and cognitive-behavioural group therapy for treatment of depressed primary-care patients, the secondary aim was to evaluate if receiving treatment according to free choice is associated with a better outcome than randomization to a particular treatment. We conducted a randomized, placebo-controlled, single-centre, 10-wk trial with five arms: sertraline (flexible dosages up to 200 mg/d) (n = 83); placebo (n = 83); manual-guided cognitive-behavioural group therapy (one individual session and nine group sessions per 90 min) (n = 61); guided self-help group (control condition, n = 59); and treatment with sertraline or cognitive-behavioural group therapy according to patients’ choice (n = 82). From 1099 consecutively screened adult patients, 368 formed the intent-to-treat population with milder forms of depression. Primary outcome was a global efficacy measure combining z-converted Hamilton Depression Rating Scale and clinician-rated Inventory for Depressive Symptomatology scores. Sertraline was superior to placebo (p = 0.03). Outcome for guided self-help groups was worse compared to cognitive-behavioural group therapy (p = 0.002) and compared to all other treatment arms including pill placebo (secondary analyses). Outcome in the patients’ choice arm was similar to that in the sertraline and cognitive-behavioural group therapy. Overall, sertraline is efficacious in primary-care patients with milder forms of depression. The superiority of cognitive-behavioural group therapy over guided self-help groups might partly be explained by ‘nocebo’ effects of the latter.

Introduction

Minor and mild to moderate depressive syndromes are common in primary care (Mulrow et al. 1998; Pincus et al. 1999; Williams et al. 1995) and associated with considerable functional morbidity, suicidality and service utilization (Fichter et al. 2008; Lyness et al. 2006; Rapaport et al. 2002; Wagner et al. 2000). Diagnostic complexity is increased by high transition rates between minor and major depression (Judd et al. 1998, 2004; Kessler et al. 1997; Lyness et al. 2006; Maier et al. 1997) and high rates of comorbidity with anxiety and somatoform disorders (Coyne et al. 1994; Sartorius et al. 1996; Wittchen & Jacobi, 2005). Evidence-based guidelines directly applicable to the heterogeneous group of
primary-care patients who are identified by primary-care providers as depressed are urgently needed.

Whereas evidence exists for the efficacy of antidepressants for more severe forms of major depression (Geddes et al. 2006; Lima & Hotopf, 2006; Mulrow et al. 2000; Schulberg et al. 1996; Silva de Lima & Hotopf, 2003), there is less evidence for patients in the milder range of depressive conditions (Markowitz, 1994; Mulrow et al. 1998; Paykel et al. 1988; Williams et al. 2000). The evidence base for the efficacy of psychotherapy with these patients is also weak (Güemes et al. 2008).

Cognitive-behavioural therapy is the most extensively investigated psychotherapy for unipolar depressive disorder (Haby et al. 2006; March et al. 2004; Parker et al. 2003; Scott, 1996). A recent meta-analysis of 11 controlled studies of the treatment of major depression and dysthymia found a moderate to large effect size of 0.77 (Haby et al. 2006). However, the patients in these studies were highly selected.

In routine care, treatment is complicated by the fact that primary-care providers must contend with depressed patients’ diverse disease concepts and treatment preferences (Dwight-Johnson et al. 2000; Hegerl et al. 2003; Jorm et al. 2000). For some patients depression is a purely psychic disorder to be treated by psychotherapy, while others have a concept of a neurobiological disorder with antidepressants as first choice treatment. It is important to know how relevant for outcome a fit is between patients’ treatment preference and the treatment that is received. Moreover, outcome may be influenced by having freely chosen a particular treatment instead of having it imposed by randomization. These facets have rarely been studied (Kendrick et al. 2006; Peveler et al. 2005; Thornett, 2001). We addressed these aspects by including a patients’ choice arm: patients who were randomized into this arm had free choice between pharmacotherapy and cognitive-behavioural group therapy (CBT).

The primary objective of this five-arm trial was to test the following hypotheses in representative, diagnostically heterogeneous and mildly depressed primary-care patients:

1. Sertraline is more effective than placebo.
2. CBT is more effective than a guided self-help group (GSG).

We addressed further secondary issues within an exploratory approach, including whether CBT was superior to pill placebo and whether giving the patients the opportunity to choose their preferred treatment (antidepressants or CBT) produced better outcomes than random assignment.

**Method**

**Design overview**

This was a prospective, randomized, single-centre, parallel-group, 10-wk clinical trial with five treatment arms. Pharmacotherapy was placebo-controlled and double-blind. Psychotherapy was evaluator-blinded. Control visits took place weekly until week 2, then in 2-wk intervals. A 1-yr follow-up was also part of the study design (results not reported here).

**Setting and participants**

We treated patients in one specialized study centre [established within the German Research Network on Depression and Suicidality (Hegerl & Pfeiffer-Gerschel, 2005)] in order to facilitate valid diagnosis. In this centre patients were only treated in the context of studies. They were referred by 18 participating primary-care providers in one German region (Nuremberg). Thus, our trial can be considered to be a single-centre study. These practices performed pre-screening using the WHO-5 Well-Being Index (Henkel et al. 2003; WHO, 1998) (cut-off score for depression: ≤13). We screened 1099 primary-care patients referred by primary-care providers to the study centre. In total, 368 patients (33.5%) met inclusion and exclusion criteria.

The eligibility criteria for the study were: a minimum age of 18 yr; subthreshold (minor) depression, dysthymia or major depressive disorder with mild to moderate severity and Hamilton Rating Scale for Depression (17-item version) (HAMD17; Hamilton, 1960) total scores ≥8 and ≤22. We excluded patients if they had current psychotherapy or antidepressants, acute suicidality, brief recurrent depression, bipolar affective disorders, addiction (alcohol, benzodiazepines, illicit drugs), schizophrenia, schizotypal personality disorder or delusional disorder, obsessive–compulsive disorder, severe somatic diseases.

The detailed study protocol was in accordance with the revised Declaration of Helsinki (World Medical Association, 1997) and Good Clinical Practice guidelines (International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use). An approval by an independent Ethics Review Committee (Medical Faculty, Ludwig-Maximilians-University Munich, Munich, Germany) was obtained, and all subjects gave written informed consent.

Psychiatrists and psychologists diagnosed patients according to DSM-IV criteria based upon a fully structured clinical interview (Composite International...
Diagnostic Interview; WHO, 1993). We applied a German computer-administered version (DIA-X) with high inter-rater reliability ($\kappa \geq 0.81$) (Wittchen & Pfister, 1997), with an added assessment of minor depression according to DSM-IV criteria.

**Randomization and interventions**

We constructed the randomization list using FoxProTM (USA). For block-stratified randomization (Schouten, 1995) four strata were accounted for: age ($\leq 65$ yr; $>65$ yr), sex, HAMD$_{17}$ total score (intensity ranges: 8–12, 13–17, 18–22), comorbidity regarding anxiety and/or somatoform disorders. Patients meeting inclusion criteria received the next patient number according to order of study entry. This number assigned patients to one of five different treatment arms (sertraline, placebo, CBT, GSG, patients’ choice). The allocation sequence was concealed to investigators and patients until treatment assignment.

**Arm 1 (sertraline) and arm 2 (placebo)**

In these arms, the visits (at weeks 1, 2, 4, 6, 8, 10) involved adjustment of medication and recording of medical history, symptoms, adverse events/severe adverse events, weight, and blood pressure. If symptom reduction was achieved, patients continued with the initial dose of the drug. Otherwise, an escalation in dose in steps of 50 mg up to 200 mg/d was possible at weeks 2, 4 and 6. Sertraline and placebo were identical in appearance. To protect the double blind, serum-level determination in the sertraline and the placebo groups was conducted after the trial. We assessed compliance by pill counts and serum-level controls: blood samples were drawn for measuring these levels at weeks 2, 4, 6, 8, 10. Patients were not informed when blood samples were drawn for measuring these levels. The results of this control procedure were not discussed with the patients at an individual level because the psychiatrists in the study centre were blinded. Pharmacotherapy was conducted by psychiatrists at the specialized study centre.

**Arm 3 (CBT) and arm 4 (GSG)**

Short-term CBT (Hautzinger, 2003) and GSG were provided for 10 wk, in nine weekly group sessions (at 90 min each) with 5–8 members after an initial individual 50-min session. CBT was offered with three manual-guided modules (planning of activities; addressing cognitive distortion; promotion of social competence), each with three sessions. The manual was derived from the cognitive theory of depression by Beck (1967). In the GSG, a supportive atmosphere was created, allowing the participants to communicate about their situation and daily life, but no psychotherapeutic intervention was allowed by the group leader.

All group sessions were videotaped. Two tapes of each 9-wk cycle were randomly selected for independent evaluation. For each session, a 15-item rating scale was used to evaluate adherence to treatment protocol of CBT or GSG with rating categories capturing distinctive elements of CBT, but also supportive, empathetic, and relationship-enhancing strategies (Hautzinger, 2003; Hautzinger et al. 1996). Criteria for an at least acceptable CBT session was 18 (of 30 possible) points. A GSG session should have no more than nine points. The tool for monitoring fidelity was a validated measure and the cut-off score for adherence was empirically derived (Hautzinger et al. 1996). Evaluation of selected tapes by three trained raters [$\kappa = 0.84$ (CBT) and $\kappa = 0.85$ (GSG)] revealed a good to excellent adherence to treatment protocols by all therapists.

**Arm 5 (patients’ choice arm)**

Patients randomized to the patients’ choice arm could select a 10-wk treatment with sertraline or CBT. Six experienced female psychotherapists provided CBT and GSG. All had several years of experience in CBT and had been additionally trained with the study manual. They conducted both CBT and GSG and received continuous supervision.

**Outcomes**

Primary efficacy measures were two standardized rating scales (for methodological aspects see Möller, 2008): the HAMD$_{17}$ (total score range: 0–52 points; cut-off score for depression: $\geq 8$; for moderate severity of depression: $\geq 15$) and the Inventory for Depressive Symptomatology (28-item version of the clinical interview: IDS; Rush et al. 1986) (range 0–84 points). The IDS was used in addition to HAMD$_{17}$ because a higher sensitivity for symptom changes in milder and atypical forms of depression has been reported (Judd et al. 2004; Rush et al. 1986). Secondary outcomes were the physician-rated Clinical Global Impressions severity score (CGI; NIMH, 1976; German version: Collegium Internationale Psychiatriae Scalarum, 1996) and response rates (see below). All HAMD$_{17}$ and IDS evaluations were performed by one of two blinded psychologists. During the course of the study, three rater training programmes were conducted. Each training session ended with a 95% rater agreement.
Statistical analysis

Confirmatory analysis for the primary efficacy measures was performed with the intent-to-treat (ITT) sample (i.e. all randomized patients). The study design combined multiple hypotheses with an adaptive interim design (Kieser et al. 1999; Lehmacher et al. 2000). The confirmatory comparisons of sertraline vs. placebo and CBT vs. GSG were considered independent. In order to maintain a global significance level of 5% for hypotheses testing of two primary outcome variables (HAMD$_9$ and IDS total scores), the rating scale scores were z-standardized and then summarized to obtain an indicator reflecting global efficacy of treatments (O’Brien, 1984). The main hypotheses were tested first for this global efficacy measure. If the respective null hypothesis could be rejected, we then tested the original main hypotheses for changes in both scales in a closed test procedure, a general method for performing more than one hypothesis test simultaneously (Lehmacher et al. 1991).

To deal with the relatively high drop-out rates and the differences between treatment arms concerning this aspect as well as concerning early vs. late drop-outs, we applied a mixed-model, repeated-measures (MMRM) analysis (Twisk, 2003). This analytical strategy allows utilization of all available data, and it is very flexible in modelling of time-effects and can handle missing data appropriately (Guergueiua & Krystal, 2004). It replaced the last observation carried forward (LOCF) approach originally designated in the study plan (the results of the LOCF analysis will be additionally reported). The model used the restricted maximum-likelihood algorithm and included fixed class effects for treatment, visit week, the interaction of treatment and visit week, as well as random effects for intercept and slope and fixed effects for covariates (sex, completer status, psychiatric comorbidity). For the analyses, all available data from post-baseline visits were utilized. We used statistical significance of the interaction of treatment and visit week to determine whether the change over time was different across the treatments.

Analyses were also performed for the per-protocol population (with sufficient compliance according to pill count and sertraline plasma levels; cf. Fig. 1).

In view of the uncertainty of power calculation assumptions, an adaptive interim analysis was planned after inclusion of 250 patients [characteristics: stopping rules; rejection of the following null hypotheses: comparative efficacy of sertraline vs. placebo; comparative efficacy of CBT vs. GSG; $\alpha = 0.05, \alpha_1 = 0.0299$ (for interim analysis) or due to futility ($\alpha_0 = 0.30$)].

Interim analysis resulted in stopping CBT and GSG because the change in the global efficacy measure demonstrated a statistically significant superiority of CBT compared to GSG [MMRM: treatment x visit interaction: $z = -3.16; p = 0.002 (< \alpha_2)$; regression coefficient ($\beta$) = -0.29, 95% confidence interval (CI) -0.47 to -0.11]. No significant differences were found concerning sertraline vs. placebo (treatment x visit interaction: $z = 1.31, p = 0.19 (> \alpha_4$, but $< \alpha_5$); $\beta = 0.12, 95\%$ CI -0.06 to 0.30).

Similarly, LOCF analysis of the change in the global efficacy measure (IDS+HAMD$_9$, z-transformed) demonstrated a statistically significant difference between CBT and GSG favouring CBT [$\Delta = -1.66, 95\%$ CI -2.78 to -0.54, $p = 0.0043 (< \alpha_4$, one-sided test] and revealed no significant differences between sertraline and placebo [$\Delta = -0.65, 95\%$ CI -1.82 to 0.51, $p = 0.27 (> \alpha_4$, but $< \alpha_5$]).

Recruitment and randomization for the remaining arms were continued with at least 30 further patients for each arm.

The statistical test between sertraline and placebo was evaluated with $p^* = p_1 \times p_2 < 0.0087$ (Fisher’s product criterion) with $p_1$ as the $p$ value for the first part of the study until interim analysis and $p_2$ as the $p$ value for the second part after interim analysis. Fisher’s product criterion is used for the combination of the $p$ values computed for the disjoint stages of a clinical trial (Wassmer, 1999). For the sake of comparability with other $p$ values, $p^*$ was transformed in $p_{global}$ by multiplication with the factor 5.75 ($= 0.05/0.0087$).

We evaluated differences in efficacy between the other treatment groups (e.g. CBT vs. placebo) in an exploratory way. Regression coefficients ($\beta$) and 95% CIs were calculated for differences between the treatment groups in efficacy based on the results of MMRM. Responders were defined as patients with a HAMD$_9$ reduction of at least 50% and HAMD$_9$ endpoint <8. $\chi^2$ analysis was used to compare the response rates between treatment groups.

We considered a $p$ value of $\leq 0.05$ statistically significant. Statistical analysis was performed by using SPSS software, version 12.0 (SPSS Inc., USA) as well as Stata version 9.0 (for MMRM analysis (Stata, 2005).

Results

Patient characteristics

A total of 1099 primary-care patients were invited to participate from May 2000 to November 2004, with 368 patients randomized and forming the ITT sample (for patient characteristics see Table 1 and Fig. 1).
Fig. 1. Allocation of patients to the five treatment arms. CBT, Cognitive-behavioural group therapy; GSG, guided self-help groups. ‘Lost to follow-up’, Patients who had neither further contact with the clinicians nor assessments of post-baseline efficacy. Therefore it is unclear whether they received treatment at all. ‘No further contact’. These patients were drop-outs with available data about post-baseline efficacy.
Efficacy

Table 2 shows the results for primary and secondary outcomes in our sample.

Sertraline vs. placebo

When combining the p value for the first and second part of the study (before and after the interim analysis) for the global efficacy measure ($p_1=0.19$, $p_2=0.03$, $p_1 \times p_2=0.006$) the obtained p value was below 0.0087 (0.006 $\Rightarrow P_{\text{global}}=0.03$). Therefore, sertraline was shown to be significantly superior to placebo in improving depressive symptoms. (This was not the case in LOCF analysis: $P_{\text{global}}=0.22$.)

According to additional analyses [HAMD$_D$ (Table 3) and IDS] sertraline and placebo differed in changes from baseline (HAMD$_D$: $z=2.19$, $p=0.03$, $\beta=0.47$, 95% CI 0.05–0.89; $\Delta=2.29$ points, Fig. 2; IDS: $z=2.26$, $p=0.02$, $\beta=0.81$, 95% CI 0.11–1.51) and response rates in the sertraline group (28/74, 38%) were higher than for placebo (16/75, 21%) ($\chi^2=4.88$, $p=0.027$, two-sided test).

Treatment differences concerning CGI severity scores were not significant ($z=1.19$, $p=0.25$, $\beta=0.05$, 95% CI $-0.03$ to 0.13).

Twelve patients of the ITT population had to be excluded from the per-protocol population because of a lack of sertraline plasma levels (sertraline or patients’ choice arms) or positive sertraline plasma levels (placebo group). Plasma levels for the other patients treated with sertraline were sufficient [sertraline group (mean $\pm$ S.D.): 18.1 $\pm$ 11.9 ng/ml ($n=51$) at week 6 and 18.9 $\pm$ 16.7 ng/ml ($n=46$) at week 10; patients’ choice sertraline group (mean $\pm$ S.D.): 20.2 $\pm$ 15.2 ng/ml ($n=37$) at week 6 and 28.7 $\pm$ 30.3 ng/ml ($n=29$) at week 10].
Table 2. Baseline values and changes of the primary and secondary outcomes in the five treatment arms

<table>
<thead>
<tr>
<th>Variables, mean (s.d.) *</th>
<th>Sertraline (n = 83)</th>
<th>Placebo (n = 83)</th>
<th>CBT (n = 61)</th>
<th>GSG (n = 59)</th>
<th>Patients' choice (n = 82)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary outcome measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAMD 17 total score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>16.4 (4.4)</td>
<td>16.1 (4.6)</td>
<td>15.9 (4.2)</td>
<td>15.6 (4.4)</td>
<td>16.1 (5.3)</td>
</tr>
<tr>
<td>Change *</td>
<td>6.8 (7.3)</td>
<td>4.5 (6.3)</td>
<td>6.7 (6.5)</td>
<td>1.9 (6.3)</td>
<td>6.1 (6.7)</td>
</tr>
<tr>
<td>[95% CI]</td>
<td>[4.9–8.6] (n = 61)</td>
<td>[2.9–6.0] (n = 65)</td>
<td>[4.7–8.8] (n = 41)</td>
<td>[−0.3 to 4.1] (n = 35)</td>
<td>[4.5–7.7] (n = 68)</td>
</tr>
<tr>
<td><strong>IDS total score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>27.5 (7.9)</td>
<td>26.5 (8.7)</td>
<td>26.6 (6.5)</td>
<td>25.8 (7.4)</td>
<td>27.1 (9.2)</td>
</tr>
<tr>
<td>Change *</td>
<td>11.6 (12.4)</td>
<td>7.5 (10.2)</td>
<td>10.9 (10.2)</td>
<td>4.3 (11.6)</td>
<td>9.5 (11.0)</td>
</tr>
<tr>
<td>[95% CI]</td>
<td>[8.2–14.9] (n = 56)</td>
<td>[4.9–10.1] (n = 62)</td>
<td>[7.6–14.2] (n = 40)</td>
<td>[0.2–8.3] (n = 34)</td>
<td>[6.7–12.2] (n = 65)</td>
</tr>
<tr>
<td><strong>Secondary outcome measure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGI, item 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>3.9 (0.7)</td>
<td>3.9 (0.6)</td>
<td>3.9 (0.8)</td>
<td>3.9 (0.7)</td>
<td>3.9 (0.7)</td>
</tr>
<tr>
<td>Change *</td>
<td>1.1 (1.5)</td>
<td>1.0 (1.2)</td>
<td>1.2 (1.2)</td>
<td>0.7 (1.2)</td>
<td>1.3 (1.3)</td>
</tr>
<tr>
<td>[95% CI]</td>
<td>[0.8–1.5] (n = 76)</td>
<td>[0.7–1.3] (n = 62)</td>
<td>[0.9–1.6] (n = 52)</td>
<td>[0.4–1.1] (n = 50)</td>
<td>[1.1–1.6] (n = 79)</td>
</tr>
</tbody>
</table>

CBT, Cognitive-behavioural group therapy; CI, confidence interval; CGI, Clinical Global Impression of severity; GSG, Guided self-help groups; HAMD 17, Hamilton Depression Rating Scale (17-item version); IDS, Inventory for Depressive Symptomatology; n, sample size; s.d., standard deviation.

* Values are based on patients for whom no data were missing.

* Change is given as difference between baseline and last visit (10th week).

Table 3. Comparison of outcome between the different treatment arms

<table>
<thead>
<tr>
<th>Between-group differences in changes of HAMD 17 sum scores [95% CI]</th>
<th>Placebo</th>
<th>CBT</th>
<th>GSG</th>
<th>Patients’ choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sertraline</td>
<td>ΔHAMD 17 = 2.29</td>
<td>ΔHAMD 17 = 0.02</td>
<td>ΔHAMD 17 = 4.87</td>
<td>ΔHAMD 17 = 0.67</td>
</tr>
<tr>
<td>Δ = 0.34 [−0.02 to 0.69]</td>
<td>z = 2.19, p = 0.03</td>
<td>z = 0.86, p = 0.39</td>
<td>z = 4.52, p &lt; 0.001</td>
<td>z = 0.94, p = 0.35</td>
</tr>
<tr>
<td>Placebo</td>
<td>ΔHAMD 17 = −2.27</td>
<td>ΔHAMD 17 = 2.58</td>
<td>ΔHAMD 17 = −1.63</td>
<td>ΔHAMD 17 = −0.30</td>
</tr>
<tr>
<td>Δ = 0.34 [−0.05 to 0.74]</td>
<td>z = 1.02, p = 0.31</td>
<td>z = -2.76, p = 0.006</td>
<td>z = 1.29, p = 0.20</td>
<td>z = 0.07, p = 0.95</td>
</tr>
<tr>
<td>CBT</td>
<td>ΔHAMD 17 = 4.85</td>
<td>ΔHAMD 17 = 0.64</td>
<td>ΔHAMD 17 = −0.67</td>
<td></td>
</tr>
<tr>
<td>Δ = 0.74 [0.28 to 1.21]</td>
<td>z = −3.34, p = 0.001</td>
<td>z = 0.09, p = 0.98</td>
<td>z = 0.78, p = 0.48</td>
<td></td>
</tr>
<tr>
<td>GSG</td>
<td>ΔHAMD 17 = 4.20</td>
<td>ΔHAMD 17 = 0.63</td>
<td>ΔHAMD 17 = −0.64</td>
<td></td>
</tr>
<tr>
<td>Δ = 0.63 [0.22 to 1.05]</td>
<td>z = 3.82, p &lt; 0.001</td>
<td>z = 0.41, p = 0.67</td>
<td>z = 0.32, p = 0.74</td>
<td></td>
</tr>
</tbody>
</table>

ΔHAMD 17: Difference (baseline – 10th week) in the sum score of the HAMD 17; CBT, cognitive-behavioural group therapy; CI, confidence interval; GSG, guided self-help groups; HAMD 17, Hamilton Depression Rating Scale (17-item version). p values based on linear mixed-effects model analyses (interaction of the factors treatment and visit week); z values indicate whether the change over time was different across the treatments. ΔHAMD 17 and Hedges’ d (bias-corrected Cohen’s d) and 95% CI based on values from patients for whom no relevant HAMD 17 data (baseline and 10th week) were missing.

The results concerning the main hypotheses are given in bold. The table has to be read as comparison between the vertical and horizontal axes. Thus, for instance, sertraline (row 1) is compared with placebo (column 1) and a significant difference is found.

Results based on interim analysis. After interim analysis, the CBT and GSG arms have been closed.

The patients’ choice arm comprises all patients who could freely choose sertraline or CBT.
Fig. 2. Changes in the mean HAMD$_{17}$ (Hamilton Depression Rating Scale – 17-item version) total scores (intent-to-treat population; observed case analysis). (a) Sertraline (—) vs. placebo (–□–). Sample sizes for weeks 0–10: sertraline $n=83$, 64, 68, 58, 52, 61; placebo $n=83$, 69, 69, 65, 55, 65. (b) Cognitive-behavioural group therapy (CBT) (—) vs. guided self-help groups (GSG) (–). Sample sizes for weeks 0–10: CBT $n=61$, 40, 46, 43, 41; GSG $n=59$, 36, 37, 37, 35. * $p<0.05$, *** $p<0.001$ (based on two-sided Mann–Whitney U tests).

The analysis of the per-protocol population (HAMD$_{17}$) showed superiority of sertraline over placebo ($F=5.53$; d.f. = 1, 101, $p=0.02$, two-sided test). In line with these results, the response rate in the sertraline group (per-protocol population) (22/48, 45.8%) was higher than in the placebo group (15/56, 26.8%) ($\chi^2=4.09$, d.f. = 1, $p=0.04$, two-sided test).

**CBT vs. GSG**

MMRM analysis revealed that CBT was superior to GSG concerning HAMD$_{17}$ and IDS total scores (treatment $\times$ visit interaction: HAMD-17: $z=-3.34$, $p=0.001$, $\beta=-0.97$, 95% CI $-1.55$ to $-0.40$; $\Delta=4.85$ points, Table 3; Fig. 2; IDS: $z=-3.36$, $p=0.001$, $\beta=-1.63$, 95% CI $-2.58$ to $-0.68$). The response rate for the CBT group (16/47, 34%) was significantly better than for the GSG (4/46, 9%) ($\chi^2=8.85$, $p=0.003$, two-sided test). However, CBT and GSG did not significantly differ in reduction in CGI severity rating (MMRM treatment $\times$ visit interaction: $z=-0.96$, $p=0.34$, $\beta=-0.05$, 95% CI $-0.16$ to 0.05). The outcome of GSG was also less favourable than that of the other three treatment arms (Table 3).

Regarding the per-protocol population, improvement tended to be greater for CBT than for GSG (HAMD$_{17}$) ($F=2.73$, d.f. = 1, 54, $p=0.10$). The response rate in the CBT group (12/33, 36.4%) was higher than in the GSG (3/24, 12.5%) ($\chi^2=4.08$, d.f. = 1, $p=0.04$, two-sided test).

**Further comparisons**

In the patients’ choice arm, 49 patients chose sertraline and 33 CBT. Overall, the patients in the patients’ choice arm did not differ in outcome from patients randomized to the sertraline or CBT arms (Table 3). This was also the case when patients choosing sertraline were compared to patients randomized to sertraline and when patients choosing CBT were compared to patients randomized to CBT. Moreover, CBT and placebo did not significantly differ in HAMD$_{17}$ outcome (Table 3).

After a median split of patients according to depression severity [HAMD$_{17}$ baseline total score $<16$ ($n=169$) vs. $>15$ points ($n=198$)], there was a statistical tendency for superiority of sertraline over placebo in patients with very mild depression (HAMD$_{17}$: $z=1.85$, $p=0.065$, $\beta=0.50$, 95% CI $-0.03$ to 1.03; $\Delta=2.49$ points) whereas CBT and GSG did not significantly differ ($z=-1.14$, $p=0.26$, $\beta=-0.36$, 95% CI $-0.99$ to 0.26; $\Delta=1.81$ points). In patients with more severe depression, the difference between CBT and GSG in efficacy – while failing to be significant due to the limited sample size – exceeded 3 points in HAMD$_{17}$ ($z=-1.40$, $p=0.16$, $\beta=-0.65$, 95% CI $-1.56$ to 0.26; $\Delta=3.26$ points), this was not true for the corresponding difference between sertraline and placebo ($z=1.46$, $p=0.14$, $\beta=0.47$, 95% CI $-0.16$ to 1.09; $\Delta=2.34$ points).

**Drop-out analysis and tolerability**

Drop-out rates were 42% (35/83) in the sertraline arm, 33% (27/83) in the placebo arm, 46% (28/61) in the CBT arm, 59% (35/59) in the GSG and 30% (25/82) in the patients’ choice arm. Differences in attrition rates between the five treatment arms were statistically significant ($\chi^2=15.07$, $p=0.005$). In the subgroup of drop-outs, the five treatment arms did not significantly differ in the rate of early drop-outs (attrition within the
first week) [sertraline: 31.4% (11/35), placebo: 33.3% (9/27), CBT: 46.4% (13/28), GSG: 40% (14/35), patients’ choice arm: 28% (7/25)]; \( \chi^2 = 2.66, p = 0.62 \).

Few patients withdrew due to adverse events: eight patients (10%) in the sertraline group, three patients (4%) in the placebo group, one patient (2%) in the CBT group (because of aggravation of leukaemia), and three patients (6%) in the patients’ choice sertraline arm.

The most frequently reported adverse events for the sertraline and sertraline-choice groups were central and peripheral nervous system disorders (e.g., tremor), gastrointestinal system disorders (e.g., nausea) and psychiatric symptoms (e.g., agitation), well-known side-effects of selective serotonin reuptake inhibitors.

Most of the adverse events were mild to moderate and dissipated during continued treatment with sertraline, without dosage adjustment.

Serious adverse events occurred in five patients during the study or within 1 month of study termination. Three serious adverse events were not considered as treatment related: two further serious adverse events were associated with sertraline: one patient discontinued sertraline because of intermittent severe diarrhoea and one suffered from a severe serotonin syndrome which improved after dose reduction.

Discussion

The present study contributes to the very limited evidence base for treatment recommendations for the large group of depressed primary-care outpatients.

Our first main hypothesis was confirmed: 10 wk sertraline treatment were superior to placebo for improving depressive symptoms as assessed with the global efficacy measure; in secondary analyses, this was also the case, e.g., regarding response rates (sertraline 38%, placebo 21%) and analysis of the per-protocol population. This finding is marked because the study was not highly powered and finding effects might be more difficult in this diagnostically heterogeneous patient group with milder forms of depression and high comorbidity. Results might reflect the fact that it was a study with only one study centre, probably avoiding variance resulting from study centre effects, and that sertraline plasma levels were monitored. The latter aspect could have increased the compliance, which could be a considerable problem in an outpatient study with primary-care patients with milder forms of depression.

The second main hypothesis was similarly confirmed: 10 wk CBT were superior to the control condition (GSG) for improvement in depressive symptoms (assessed with the global efficacy measure). This difference was already significant at the interim analysis. Therefore, according to protocol, these arms were closed earlier. This effect was consistently supported by secondary analyses: CBT was also superior to GSG in terms of response rates (CBT 34%, GSG 9%) and – at least by trend – in per-protocol analysis. However, interpretation of these results is less straightforward than for the sertraline-placebo comparison. GSG were significantly less effective than CBT, and also less effective than all other arms, including placebo. This unexpected finding suggests that the GSG were not a ‘psychotherapy placebo’ condition, analogously to the pill placebo group. Because blinding concerning CBT vs. GSG was impossible, patients knew that they were only in the control group. This could have resulted in frustration and the observed high drop-out rate (59%). In addition, the therapists in the GSG were confronted with the situation that they were not allowed to make specific interventions or flexibly apply their therapeutic skills, but had to maintain a more passive role in the sessions. This factor may have negatively affected the patient-therapist interaction.

The finding that the psychotherapy control group had a worse outcome than the pill placebo group draws attention to the more general problem of interpreting results in controlled psychotherapy research (Klein, 1996). It is extremely difficult to devise an adequate psychotherapy ‘placebo’ condition. The reasons include uncertainty about the active ingredients in psychotherapy, the inability to blind therapists and patients, and the difficulty of devising a placebo condition that is credible to patients. Because blinding is not possible, superiority of psychotherapeutic interventions relative to control conditions such as waiting list or non-specific support might reflect not the efficacy of the psychotherapy, but only the negative effects of being assigned to the control condition (‘nocebo effect’). Direct comparisons of psychotherapy controls and pill placebo controls are rarely reported, even where data are available. However, a possible negative effect of waiting-list control groups is suggested by a meta-analysis (Robinson et al. 1990) that showed the mean effect size of psychotherapy is much higher (0.84) when compared to a waiting list (29 studies) than when compared to drug placebo (effect size 0.28, nine studies). Cognitive-behavioural therapy does not appear to be superior to pill placebo for depression in a number of trials (DeRubeis et al. 2005; Dimidjian et al. 2006; Elkin et al. 1989), with the one exception being a trial with patients with atypical depression (Jarrett et al. 1999).
In the present study, when comparing CBT to drug placebo, no significant differences for HAMD$_{17}$ and response rates were found despite similar antidepressant effects for CBT and sertraline. This might be because the statistical power for this comparison was reduced due to CBT being stopped after the interim analysis. The reduction in HAMD$_{17}$ observed in the CBT group was virtually identical to that in the sertraline group (6.7 vs. 6.8 points). This reduction is clearly above what has been proposed as threshold for clinically relevant antidepressant effects [at least 3 points in the HAMD$_{17}$ scale (NICE, 2004)] and is significant for quality of life.

Patients randomized into the fifth arm (patients’ choice arm), who could choose their preferred treatment (sertraline or CBT), did not differ in their outcome from patients randomized in the sertraline or CBT arm. It appears that giving a patient the possibility to choose and receive the preferred treatment did not strongly affect outcome. Ward et al. (2000) also found no outcome differences in studies comparing depressed patients receiving a drug or psychotherapeutic treatment according to randomization or patients’ preference. However, a randomized, controlled trial to compare the cost-effectiveness of tricyclic antidepressants, selective serotonin reuptake inhibitors and lofepramine in 327 patients with a new episode of depressive disorders (Kendrick et al. 2006; Peveler et al. 2005) revealed that having a preference considering initial treatment was subsequently associated with a lower rate of switching therapy to another antidepressant.

We also addressed the question whether our positive main findings could be explained by a difference that is limited to the more severely depressed part of the sample (with baseline HAMD$_{17}$ total scores ≥ 16 points). This did not seem to be the case for sertraline because the sertraline-placebo differences in efficacy were even more pronounced in patients with very mild depression (ΔHAMD$_{17}$: 2.49 points) than in patients with more severe depression (2.34 points). This finding has to be interpreted with caution due to the limited power, but is in accord with the main result of a study (Khan et al. 2007) showing that more severe depressive symptoms before randomization were not significantly associated with an increased placebo-verum difference. Thus, treatment with sertraline may not only be an option in patients with mild to moderate major depression, but also in patients with very mild depression.

Outcome differences were found between sertraline and placebo as well as between CBT and GSG for the HAMD$_{17}$ (and IDS), but not the CGI. Differences in sensitivity of the two measures as well as the fact that HAMD$_{17}$ and IDS were performed by blinded independent raters and CGI by psychiatrists in the study centre might explain this discrepancy.

Among the limitations of our study is the moderate sample size causing slight structural differences in the five treatment arms and the relatively small statistical power concerning the comparison between the randomized and patients’ choice arms. In hindsight, stopping CBT due to the finding of differences with GSG in planned interim analyses resulted in low power for exploratory comparisons between sertraline and CBT.

We intended to investigate a quite representative sample of depressed primary-care patients by (1) reliance on direct primary-care providers’ referrals and screening of primary-care patients, (2) acceptance of patients with depression not otherwise specified, with minor depression and with comorbid anxiety, and (3) the entry criterion of a HAMD$_{17}$ total score >7, but <23. Eight points in the HAMD$_{17}$ was chosen as lowest entry criterion because lower scores reflect absence of depression (APA, 2000). Nevertheless, because of ethical reasons we had to exclude patients with other comorbid conditions that are frequently seen in primary care (e.g. suicidality and substance abuse).

The high drop-out rates are another limitation. This is probably due to only outpatients with mild symptomatology having been recruited, for whom the relative burden of tolerating symptoms vs. investing in treatment might be less. Moreover, the CBT fidelity assessment could have been strengthened by coding therapist competence in addition to adherence to the model.

What can now be recommended to primary-care providers? Sertraline was found to be effective even in this heterogeneous group of outpatients with mild-to-moderate forms of depressive syndromes. The observed sertraline-placebo difference (2.3 points on HAMD$_{17}$, effect size (Hedges’ $d$): 0.34, 95% CI – 0.02 to 0.69) is below what has been proposed as the threshold of clinical significance (placebo-verum difference of ≥ 3 points in HAMD$_{17}$; NICE, 2004). However, this placebo-verum difference is likely to be a gross underestimation of the differences between receiving a specific antidepressant treatment vs. ‘watchful waiting’ in daily practice. One reason for that is that the placebo effects generated by randomized controlled studies can be expected to be much larger than positive effects by receiving unspecific support within a watchful waiting strategy. This point is highlighted by the fact that the difference between sertraline and the GSG was 4.9 points in HAMD$_{17}$, which is more than
double the placebo-sertraline difference and a clearly clinically relevant effect. The situation of patients in GSG might be partly comparable to that of those in the ‘watchful waiting’ group. To use placebo-verum differences for the evaluation of the clinical relevance of a treatment with antidepressants has the risk to be grossly misleading, especially for patients with milder forms of depression. Efficient treatments might be erroneously discarded, when using placebo-verum differences of ≥ 3 points in HAM-D17 as a criterion for clinical relevance. Therefore, sertraline treatment is worthy of consideration in the context of a close monitoring of the individual cost benefits.

It is particularly relevant to routine practice that results were obtained with a diagnostically heterogeneous group of depressed patients because a reliable differential diagnosis of the broad spectrum of milder and minor depressive disorders might not be achievable at the primary-care level in most healthcare systems. Concerning CBT, no strong conclusions concerning efficacy can be drawn from the present study. The possibility of ‘nocebo’ effects in psychotherapy studies using a ‘psychotherapy control’ condition is a general problem in psychotherapy research making definitive statements about efficacy difficult. A promising strategy might be to integrate elements of CBT like problem-solving treatment in a tailored collaborative care model for depressed patients in primary care since this approach could be shown to be significantly more effective than conventional care for depression in a broad range of primary-care practices (Gallo et al. 2007; Hunkeler et al. 2006; Katon et al. 1999; Unützer et al. 2002).

Acknowledgements

This project was supported by the German Ministry for Education and Research within the promotional emphasis ‘German Research Network on Depression and Suicidality’. The funding agency had no rule in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript. We thank the participating patients, general practitioners for motivating patients to visit the study centre, especially Veit Wambach, M.D., and Vanadis Kamm-Kohl, M.D., and practice staff. We are indebted to Michaela König, M.Sc., Stephanie Lösch, M.Sc., and Matthias Stürmer, M.Sc. for monitoring the progress of the study; Evelyn Poth, M.Sc. for administrative support; Ute Hägele, M.D., and Patrick Bussfeld, M.D. for recruitment of the patients; Markus J. Schwarz, M.D. for measurement of sertraline plasma levels, Professor Hans-Jürgen Möller, M.D. for contributions to design and methods, Rico Hylla, M.Sc. and Herbert Matschinger, Ph.D. for statistical advice. [Trial registration no: NCT00226642 (http://www.clinicaltrials.gov).]

Statement of Interest

Professor Hegerl has been on or is participating in advisory boards of Lilly, Wyeth, Lundbeck and Sanofi-Aventis. He has received financial support for an Investigator Initiated Trial from Lundbeck. He has received honoraria for single lectures as well as sponsorship for an internet service from different pharmaceutical companies, health insurances and other parties.

References


