The clinical effectiveness of CBT-based guided self-help interventions for anxiety and depressive disorders: a systematic review

G. Coull* and P. G. Morris

Clinical and Health Psychology, University of Edinburgh, UK

Background. Cognitive-behavioural therapy (CBT)-based guided self-help (GSH) has been suggested to be an effective intervention for mild to moderate anxiety and depression, yet the evidence seems inconclusive, with some studies reporting that GSH is effective and others finding that GSH is ineffective. GSH differs in important respects from other levels of self-help, yet the literature regarding exclusively guided self-help interventions for anxiety and depression has not been reviewed systematically.

Method. A literature search for randomized controlled trials (RCTs) examining CBT-based GSH interventions for anxiety and depressive disorders was conducted. Multiple electronic databases were searched; several journals spanning key disciplines were hand-searched; reference lists of included review articles were scanned and relevant first authors were contacted.

Results. Thirteen studies met the inclusion criteria. Meta-analysis indicated the effectiveness of GSH at post-treatment, although GSH was found to have limited effectiveness at follow-up or among more clinically representative samples. Studies that reported greater effectiveness of GSH tended to be of lower methodological quality and generally involved participants who were self-selected rather than recruited through clinical referrals.

Conclusions. Although there is support for the effectiveness of CBT-based GSH among media-recruited individuals, the finding that the reviewed RCTs had limited effectiveness within routine clinical practice demonstrates that the evidence is not conclusive. Further rigorous evidence based on clinical populations that examines longer-term outcomes is required before CBT-based GSH interventions can be deemed effective for adults accessing primary care services for treatment of anxiety and depression.

Received 7 June 2010; Revised 4 February 2011; Accepted 7 February 2011; First published online 15 June 2011

Key words: CBT, clinical practice, effectiveness, review, self-help.

Introduction

There has been a recent impetus in the UK to improve patients’ access to psychological therapies (DoH, 2005). This has been targeted through a stepped care model in which the intensity of intervention is matched to the severity of mental health symptoms. Stepped care has the potential to maximize clinical benefits from available therapeutic resources (Bower & Gilbody, 2005). National Institute for Clinical Excellence (NICE) guidelines recommend the provision of cognitive-behavioural therapy (CBT)-based guided self-help (GSH) intervention for anxiety and depressive disorders as part of the stepped care approach (NICE, 2007, 2009). Despite national recommendations advocating GSH, the evidence is inconclusive and a systematic review of exclusively guided self-help interventions for anxiety and depressive disorders has not been conducted.

GSH can be regarded as a slightly more intensive treatment than ‘pure’ self-help, in that it involves the support of a health professional to ‘guide’ the patient in the use of a self-help intervention or ‘health technology’ (e.g. a written manual or website). Thus, a key difference between GSH and non-GSH interventions is the presence of therapist input and the potential impact of therapist factors upon GSH effectiveness outcomes. There is considerable variability within GSH interventions in terms of: the experience and type of professional providing the guidance; the quantity of input provided; and the nature of the health technology being advocated. Although effectiveness for GSH interventions for depression has

* Address for correspondence: Dr G. Coull, Clinical and Health Psychology, School of Health in Social Science, University of Edinburgh, Edinburgh EH8 9AG, UK.
(Email: g.j.coull@sms.ed.ac.uk)
been indicated in some instances (e.g. Gellatly et al. 2007), the evidence for effectiveness within clinical research trials or routine primary care services varies considerably (Khan et al. 2007). For instance, Lucock et al. (2008) describe controlled studies of GSH that have not demonstrated clinical benefits and highlight the minimal number of well-designed controlled studies of GSH and Lovell et al. (2008) convey the lack of consensus regarding the optimal format and provision of GSH. These conclusions, in addition to a tendency within research literature for a blurred demarcation between the concepts of GSH and non-GSH interventions, indicate the importance in specifically reviewing the clinical effectiveness of guided self-help for anxiety and depressive disorders.

Systematic reviews of research examining self-help interventions for anxiety and depressive disorders indicate their effectiveness (e.g. Bower et al. 2001; Morgan & Jorm, 2008), but temper their conclusions because of the heterogeneous mix of self-help interventions reviewed. Other reviews within the area have either not been systematic (e.g. Newman et al. 2003), have not distinguished between ‘pure’ self-help and GSH (e.g. den Boer et al. 2004), or have reviewed a combination of both self-help and GSH interventions (e.g. Gellatly et al. 2007). Given (i) the ambiguity surrounding the effectiveness of GSH interventions (particularly in the longer term), (ii) the inherent differences between GSH and non-guided (‘pure’) self-help, and (iii) the absence of a systematic review exclusively examining the effectiveness of GSH interventions for anxiety and depression, the aim of this review was to systematically evaluate the clinical effectiveness of guided self-help interventions for anxiety and depressive disorders.

Method

Reporting within this systematic review followed guidance as outlined by the Centre for Reviews and Dissemination (CRD), The University of York (www.york.ac.uk/inst/crd/), which forms part of the National Institute for Health Research and produces internationally accepted guidelines for undertaking systematic reviews.

Inclusion and exclusion criteria

Study design

Studies were eligible for inclusion if they reported randomized controlled trials (RCTs) that examined GSH interventions in comparison to either: ‘pure’ self-help (i.e. interventions without therapist contact); usual psychological treatment (e.g. standard CBT); or waiting list control conditions.

Population

Included studies were based solely on adult participants (within the age range of 17–64 years) with anxiety or depressive disorders, regardless of gender, race or nationality. The presence of anxiety or depressive disorder was based upon either a structured clinical interview for assessment of a diagnosis according to DSM-IV or ICD-10 criteria, or indicated by validated assessment scales adopting cut-off scores to establish clinically significant symptomatology: that is, \( \geq 11 \) on the anxiety scale of the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983); \( \geq 3 \) on the General Health Questionnaire (GHQ; Goldberg & Williams, 1988); \( \geq 16 \) on the Center for Epidemiologic Studies Depression Scale (CES-D; Bouma et al. 1995); or \( \geq 14 \) on the Beck Depression Inventory II (BDI-II; Beck et al. 1996).

Anxiety disorders included within this review are: panic disorder (with or without agoraphobia); generalized anxiety disorder; obsessive-compulsive disorder; social anxiety/phobia; phobias; and mixed anxiety disorder samples. Major depressive disorder populations were included in this review but sub-threshold clinical depression and dysthymia were excluded.

Interventions

Definitions of GSH vary between studies; Lovell et al. (2008) refer to GSH as ‘involving a CBT-based self-help resource and limited support from a healthcare professional’ whereas Mead et al. (2005) describe the GSH model as an example of minimal contact where the focus is on self-help, but the therapist teaches effective use of the self-help resource. GSH can be provided either by professionals (i.e. therapists with a postgraduate mental health qualification) or by para/non-professionals (i.e. therapists without a postgraduate mental health qualification). Inclusion of the latter group within this review is harmonious with the findings of a Cochrane review that indicated no difference between professionals and paraprofessionals in effecting change within treatment outcomes of individuals with anxiety and depressive disorders (Boer et al. 2005).

Within the present review, GSH is defined as an individual’s access to CBT-based self-help materials (e.g. books/manuals/internet) in the treatment of mild to moderate anxiety or depressive disorders, guided by the active support of a professional or paraprofessional therapist for no less than 30 min and no more than 3 h in total. Studies in which therapist support consisted solely of reminders or assessment monitoring were excluded, as were studies that had less than a 1-month follow-up evaluation. Studies
without an appropriate control condition or with un-interpretable findings were also excluded.

Outcome measures

Studies assessing clinical effectiveness health outcomes through validated observer and/or self-report measurement tools of anxiety and depression were eligible for inclusion. If effect sizes for primary outcome measures comparing treatment and control groups at post-treatment and follow-up were not documented, they were calculated using the formula for Cohen’s $d$: \[
\frac{(treatment\ mean - control\ mean)}{pooled\ standard\ deviation}
\]

Literature search strategies

Searches were limited to studies published in English because of lack of feasibility for translation of texts. The literature search was initially conducted in July 2009. The Cochrane Database of Abstracts of Reviews of Effects (DARE) was searched to verify that a similar review had not been conducted recently. To ensure this initial search was as comprehensive as possible, DARE was searched using the more inclusive term: ‘self-help’ in addition to ‘guided self-help’ and ‘depressi*’ OR ‘anxiety’. This search revealed only two articles loosely pertinent to the current review: first, a Cochrane protocol (i.e. not a review) of brief media-delivered interventions for psychological problems (Mayo-Wilson & Montgomery, 2007); and second, a systematic review of randomized and non-randomized trials of self-help, that is, not solely RCTs and not exclusively examining guided self-help (Bower et al. 2001).

Subsequently, screening of texts was conducted by searching the following electronic databases: PsycINFO (1990–2009); CINAHL (1990–2009); EMBASE (1990–2009); and Medline (1990–2009). Searches were conducted within the domains of title, abstract and keywords. The following search string was used within each database: (‘guided self-help’ OR ‘assisted self-help’ OR ‘facilitated self-help’ OR ‘supervised self-help’ OR ‘supported self-help’ OR ‘minimal intervention*’ OR ‘minimal contact’) AND (‘anxiety’ OR ‘depressi*’). These four databases were searched again using the same search string in May 2010 to account for any relevant articles published in the duration since July 2009 when the original literature search had been conducted.

Thereafter, to reduce any effect of publication bias, G.C. contacted the primary authors of included studies and key review articles (e.g. Bower et al. 2001; Gellatly et al. 2007) to incorporate any unpublished studies that might meet inclusion criteria. Twenty-two authors were approached, of whom three could not be contacted and two did not respond. The 17 responding authors suggested 18 articles (both published and unpublished), but none of these met inclusion criteria for the current review. Additionally, relevant journals within the years 2006–2009 were hand searched: British Journal of General Practice, British Journal of Psychiatry and Psychological Medicine. The search process (as detailed in Table 1) was completed by a manual search of each reference list from the included

### Table 1. Summary of literature sources and resultant review articles

<table>
<thead>
<tr>
<th>Source of articles</th>
<th>Number of potentially relevant articles initially screened for inclusion</th>
<th>Number of articles included within this review</th>
<th>Review article number*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CENTRAL</td>
<td>34</td>
<td>3</td>
<td>8, 9, 12</td>
</tr>
<tr>
<td>PsycINFO</td>
<td>67</td>
<td>4</td>
<td>6, 8, 9, 11</td>
</tr>
<tr>
<td>EMBASE</td>
<td>79</td>
<td>5</td>
<td>6, 8, 9, 11, 12</td>
</tr>
<tr>
<td>Medline</td>
<td>82</td>
<td>4</td>
<td>6, 8, 9, 12</td>
</tr>
<tr>
<td>CINAHL</td>
<td>45</td>
<td>3</td>
<td>8, 9, 12</td>
</tr>
<tr>
<td>Suggested papers after contacting relevant first authors</td>
<td>18</td>
<td>2</td>
<td>6, 13</td>
</tr>
<tr>
<td>Hand-searching of relevant journals (2006–2009)</td>
<td>9</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>British Journal of General Practice</td>
<td>5</td>
<td>2</td>
<td>4, 5</td>
</tr>
<tr>
<td>British Journal of Psychiatry</td>
<td>11</td>
<td>2</td>
<td>8, 10</td>
</tr>
<tr>
<td>Psychological Medicine</td>
<td>428</td>
<td>4</td>
<td>1, 2, 3, 7</td>
</tr>
<tr>
<td>Manual search of reference list from included review articles</td>
<td>778</td>
<td>13</td>
<td>1–13</td>
</tr>
</tbody>
</table>

* Review article numbers denote articles as follows: 1: Abramowitz et al. (2009); 2: Andersson et al. (2005); 3: Carlbring et al. (2006); 4: Carlbring et al. (2007); 5: Furmark et al. (2009); 6: Lovell et al. (2008); 7: Marks et al. (2004); 8: Mead et al. (2005); 9: Richards et al. (2003); 10: Salkovskis et al. (2006); 11: Schneider et al. (2005); 12: van Boeijen et al. (2005); 13: Warmerdam et al. (2008).
articles within this review, resulting in a total sample comprising 778 studies.

The titles and abstracts of the 778 potentially relevant studies were screened for initial assessment of their suitability according to inclusion and exclusion criteria, resulting in 41 studies. Upon further detailed reviewing of these studies, 28 studies were excluded for reasons outlined in Appendix 1. The final review was based on the remaining 13 studies. The flow of the literature review process is illustrated in Fig. 1.

Assessment of quality of included studies

A recent Cochrane protocol (Mayo-Wilson & Montgomery, 2007) for media-delivered CBT for anxiety disorders in adults concluded that ‘existing scales for measuring the quality of controlled trials have not been properly developed, are not well-validated and can give differing ratings of trial quality in systematic reviews’. They advocate the a priori identification of relevant quality criteria that are pertinent to the specific review being conducted. The CRD recommends that quality criteria should encompass an assessment of: the risk of bias; the choice of outcome measure; statistical issues; quality of reporting; quality of the intervention; and external validity (www.york.ac.uk/inst/crd/). Extending from these themes and given consideration of the review topic, the current review encompasses a checklist of 10 quality criteria identified a priori, which are outlined in Table 3. The 10 quality criteria were assessed in accordance with six outcome ratings as used by the Scottish Intercollegiate Guidance Network (SIGN) for assessing the methodological quality of RCTs. G.C. classified each quality criterion for each study in terms of one of the following six outcome ratings: ‘well-covered’ (2 points); ‘adequately addressed’ (1 point); and ‘poorly addressed’, ‘not addressed’, ‘not reported’ and ‘not applicable’ (all 0 points). P.G.M. independently reviewed the quality of nine of the 13 review articles, producing exact agreement on 78% (70/90) of methodological quality ratings; we differed by one point (e.g. well-covered versus adequately addressed) on 20% (18/90) of items and by two points (e.g. well-covered versus poorly addressed) on 2% (2/90) of items. All criteria with differences between raters were reviewed and amended where appropriate.

Results

Characteristics of included studies

The 13 studies identified for the review were all RCTs. Seven studies evaluated the effects of GSH upon anxiety disorders, four studies focused exclusively
upon depression and two studies considered both anxiety and depression. Effect size calculations at pretreatment indicated no differences between treatment and control groups in terms of primary outcome measures. Details of study characteristics and key findings are presented in Table 2.

Quality of included studies

Table 3 provides ratings for each of the studies on the 10 quality criteria. Although the rating scale adopted does not provide an exact comparative measure across studies, it offers a guide to their relative methodological strengths. It suggests that Mead et al. (2005) and Salkovskis et al. (2006) conducted the methodologically strongest studies, although the majority of reviewed studies were of average quality overall.

As only four studies (Marks et al. 2004; Mead et al. 2005; Salkovskis et al. 2006; Warmerdam et al. 2008) explicitly reported details regarding the validity or reliability of their outcome measures, we independently examined the psychometric properties for all primary outcome measures outlined across the review articles. All measures were found to be valid and reliable for the relevant populations. In terms of the statistical variables (i.e. quality criteria: vi, vii and viii), one study seemed to be particularly robust (Salkovskis et al. 2006). This study and those by Andersson et al. (2005), Mead et al. (2005) and Schneider et al. (2005) were the only ones to be sufficiently powered. The degree of treatment fidelity applied to interventions was not reported for the majority of studies, although Mead et al. (2005) and Lovell et al. (2008) considered the impact of such integrity upon effectiveness outcomes.

Six studies (Marks et al. 2004; Andersson et al. 2005; Carlbring et al. 2006, 2007; Abramowitz et al. 2009; Furmark et al. 2009) reported large effect sizes demonstrating effectiveness for GSH relative to controls at post-treatment. However, most of these studies were based upon media-recruited samples rather than samples recruited by mental health professionals and only one was sufficiently powered. Furthermore, the effectiveness of GSH relative to controls for these studies was typically either not reported at longer-term follow-up (Table 2) or indicated only a small effect size at follow-up (Furmark et al. 2009). By contrast, the studies that scored more highly on the methodological quality criteria (see the overall quality scores in Table 3) tended to be based on clinical samples and mostly demonstrated limited or no effectiveness of GSH compared to controls, particularly at longer-term follow-up: effect sizes of 0.18 and 0.03 were reported by Mead et al. (2005) and Salkovskis et al. (2006) respectively. The methodologically strongest RCTs indicated that GSH did not lead to improved mental health outcomes in the longer term (e.g. ≥3 months) with respect to waitlist control or general practitioner (GP) usual care (Mead et al. 2005; Salkovskis et al. 2006).

Meta-analysis

Meta-analysis was conducted on 11 of the 13 reviewed studies reporting data post-intervention (Lovell et al. 2008 and Mead et al. 2005 did not report post-treatment data). Where studies reported more than one primary outcome measure, we chose the first reported primary measure to ensure that no study was over-represented in the meta-analysis. Findings at post-treatment indicated a mean-weighted effect size of 0.69, suggesting considerable effectiveness of GSH compared to control conditions at post-treatment. However, seven of these 11 studies recruited participants primarily through the media rather than clinical settings, with a mean effect size for media-recruited studies of 1.02, compared to a mean effect size for more clinically representative studies of 0.31. The Q-test of homogeneity revealed significant heterogeneity among effect sizes ($Q = 29.13, df = 10, p < 0.01$), indicating greater variation than would be expected on the basis of sampling variability. Although further exploration of this heterogeneity and the potential effects of recruitment method would have been useful, the small number of studies prohibited further detailed analysis.

Meta-analysis of effect sizes relating to differences between intervention and control groups was also conducted at follow-up and was feasible for nine of the 13 studies. The mean weighted effect size at follow-up of 0.32 was further reduced to 0.19 after excluding one study (Warmerdam et al. 2008) that had a low methodological rating and seemed to exert undue influence on the analysis. The Q test of homogeneity at follow-up indicated no significant heterogeneity ($Q = 10.45, df = 8, p = 0.3$).

Discussion

This systematic review and meta-analysis conveys mixed findings for the effectiveness of GSH treatment for anxiety and depressive disorders. Although GSH seems to be significantly more effective than waitlist control conditions if we consider only outcomes immediately post-treatment among studies that recruited participants primarily through media advertisement, this effectiveness is considerably diminished among clinically representative samples or at follow-up. The evidenced heterogeneity at post-treatment and apparent differences according to the recruitment method
Table 2. Characteristics, effect sizes and key finding of reviewed studies

<table>
<thead>
<tr>
<th>Study Country</th>
<th>Diagnosis</th>
<th>Gender (% female)</th>
<th>Mean age (years) at baseline (S.D.)</th>
<th>Intervention arms (n) at follow-up</th>
<th>Amount of therapist guidance (min)</th>
<th>Primary outcome measure</th>
<th>Follow-up period (months)</th>
<th>Recruitment method</th>
<th>Post-treatment effect size (weighted)</th>
<th>Follow-up effect size (weighted)</th>
<th>Key finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abramowitz et al. (2009) USA</td>
<td>Mild–moderate social phobia</td>
<td>76</td>
<td>43.4 (10.8)</td>
<td>GSH (11) Waitlist control (N.A.)</td>
<td>180</td>
<td>BSPS</td>
<td>3</td>
<td>Media/clinical</td>
<td>1.12 (6.45)</td>
<td>N.A.</td>
<td>Self-help with minimal contact (therapist visit) group significantly superior to waitlist control group in terms of social anxiety symptoms at post-treatment.</td>
</tr>
<tr>
<td>Andersson et al. (2005) Sweden</td>
<td>Mild–moderate depression</td>
<td>78</td>
<td>36.4 (11.5)</td>
<td>Internet CBT + emails + discussion group (36) Discussion group control (N.A.)</td>
<td>60–90</td>
<td>BDI</td>
<td>6</td>
<td>Media</td>
<td>0.85 (14.93)</td>
<td>0.00 (0.00)</td>
<td>Internet-based therapy with minimal therapist contact group significantly superior to discussion waitlist control group at post-treatment.</td>
</tr>
<tr>
<td>Carlbring et al. (2006) Sweden</td>
<td>Mild–moderate panic disorder</td>
<td>60</td>
<td>36.7 (10.0)</td>
<td>Internet-guided CBT + weekly telephone calls (26) Waitlist control (N.A.)</td>
<td>120</td>
<td>ACQ</td>
<td>9</td>
<td>Media</td>
<td>1.52 (34.18)</td>
<td>N.A.</td>
<td>Internet-based self-help with minimal contact group significantly improved compared to waitlist control at post-treatment.</td>
</tr>
<tr>
<td>Carlbring et al. (2007) Sweden</td>
<td>Mild–moderate social phobia</td>
<td>59</td>
<td>32.4 (9.1)</td>
<td>Internet-guided CBT + weekly telephone calls (27) Waitlist control (N.A.)</td>
<td>95</td>
<td>LSAS</td>
<td>12</td>
<td>Media</td>
<td>0.98 (15.37)</td>
<td>N.A.</td>
<td>Compared to waitlist control group, the internet-GSH group demonstrated significant improvement in social anxiety measures at post-treatment.</td>
</tr>
<tr>
<td>Furmark et al. (2009) Sweden</td>
<td>Social anxiety disorder</td>
<td>78</td>
<td>35.0 (10.2)</td>
<td>Internet-guided CBT + weekly emails (36) Waitlist control (33)</td>
<td>135</td>
<td>LSAS</td>
<td>12</td>
<td>Media</td>
<td>0.78 (12.57)</td>
<td>0.25 (4.34)</td>
<td>Internet-GSH led to significant improvements in social anxiety compared to waitlist control group at post-treatment.</td>
</tr>
<tr>
<td>Lovell et al. (2008) UK</td>
<td>Mild–moderate (severe) depression</td>
<td>74</td>
<td>37.6 (12.4)</td>
<td>GSH (coaches) (19) GP treatment as usual (23)</td>
<td>132</td>
<td>BDI</td>
<td>3</td>
<td>Clinical</td>
<td>N.A.</td>
<td>0.18 (2.00)</td>
<td>No significant difference between GSH group and usual care group in terms of BDI scores at 3-month follow-up.</td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Diagnosis</td>
<td>Sample Size</td>
<td>Mean Age (SD)</td>
<td>Treatment Group 1</td>
<td>Treatment Group 2</td>
<td>Follow-Up</td>
<td>Conclusion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------</td>
<td>----------------------------------</td>
<td>-------------</td>
<td>---------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>-----------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marks et al. (2004) UK</td>
<td></td>
<td>Phobia and panic disorder</td>
<td>69</td>
<td>38.0 (12.0)</td>
<td>Internet-guided +</td>
<td>Relaxation</td>
<td>120</td>
<td>Media/clinical 1.38 (15.55) N.A. Minimal contact internet-guided self-exposure was as effective as clinician-guided exposure therapy; both demonstrated significant change compared to relaxation control group at post-treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mead et al. (2005) UK</td>
<td></td>
<td>Mild–moderate (severe) anxiety depression</td>
<td>72</td>
<td>38.7 (10.7)</td>
<td>Assistant GSH (50)</td>
<td>Waitlist control (53)</td>
<td>60–120</td>
<td>HADS 3 Clinical 0.18 (4.50) No significant difference between GSH group and waitlist control group on HADS at 3-month follow-up</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Richards et al. (2003) UK</td>
<td></td>
<td>Mild–moderate anxiety depression</td>
<td>84</td>
<td>39.2 (12.6)</td>
<td>Nurse-guided CBT self-help (20)</td>
<td>GP treatment as usual (21)</td>
<td>120–180</td>
<td>CORE 3 Clinical 0.49 (4.54) 0.31 (3.03) Nurse GSH was not significantly more effective in terms of primary outcomes than GP usual care at 3-month follow-up</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salkovskis et al. (2006) UK</td>
<td></td>
<td>Moderate (severe) depression</td>
<td>78</td>
<td>39.2 (13.3)</td>
<td>GP GSH (38)</td>
<td>GP treatment as usual (39)</td>
<td>120–180</td>
<td>BDI 6 Clinical 0.14 (0.41) 0.03 (0.62) No significant differences between GSH and GP usual care at 6-month follow-up</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schneider et al. (2005) UK</td>
<td></td>
<td>Phobias and panic disorder</td>
<td>74</td>
<td>39.0 (11.0)</td>
<td>Internet-guided self-exposure (31)</td>
<td>Internet-guided minimal CBT (13)</td>
<td>115</td>
<td>Main problem 1 Clinical 0.10 (0.11) 0.39 (3.81) At 1-month follow-up, improvement was significantly greater if self-help included exposure instructions versus minimal CBT excluding exposure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>van Boeijen et al. (2005)</td>
<td></td>
<td>Anxiety (GAD) panic</td>
<td>62</td>
<td>38.8 (12.7)</td>
<td>GP GSH (53)</td>
<td>GP guidelines (26)</td>
<td>100</td>
<td>STAI 9 Clinical 0.51 (4.52) 0.33 (5.73) GSH produced as much improvement as less structured GP guidance in terms of anxiety outcomes at 9-month follow-up</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warmerdam et al. (2008)</td>
<td></td>
<td>Mild–moderate depression</td>
<td>71</td>
<td>45.0 (12.1)</td>
<td>Internet/email-guided CBT (88)</td>
<td>Waitlist control (87)</td>
<td>160</td>
<td>CES-D 1 Media 0.54 (11.39) 0.69 (30.66) Internet-GSH was effective in significantly reducing depressive symptoms at 1-month follow-up compared to waitlist control group</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GSH, Guided self-help; CBT, cognitive-behavioural therapy; GP, general practitioner; GAD, generalized anxiety disorder; N.A., not applicable; S.D., standard deviation; ACQ, Agoraphobic Cognitions Questionnaire; BDI, Beck Depression Inventory; BSFS, Brief Social Phobia Scale; CES-D, Center for Epidemiologic Studies Depression Scale; CORE, Clinical Outcomes in Routine Evaluation; HADS, Hospital Anxiety and Depression Scale; LSAS, Liebowitz Social Anxiety Scale; STAI, State–Trait Anxiety Inventory.
Table 3. Ratings of study quality for included studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Randomization</th>
<th>Allocation</th>
<th>Baseline assessed</th>
<th>Confounds controlled</th>
<th>Outcome measures</th>
<th>Attrition to-treat</th>
<th>Intention-to-treat</th>
<th>Power</th>
<th>Fidelity</th>
<th>Generalizability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abramowitz et al. (2009)</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Well-covered</td>
<td>Adequately addressed</td>
<td>Not reported</td>
<td>Well-covered</td>
<td>Not reported</td>
<td>Adequately addressed</td>
<td>Poorly addressed</td>
<td>4.0</td>
</tr>
<tr>
<td>Andersson et al. (2005)</td>
<td>Adequately addressed</td>
<td>Not reported</td>
<td>Poorly addressed</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>Well-covered</td>
<td>Not reported</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>5.0</td>
</tr>
<tr>
<td>Carlbring et al. (2006)</td>
<td>Well-covered</td>
<td>Not reported</td>
<td>Well-covered</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>Well-covered</td>
<td>Not reported</td>
<td>Adequately addressed</td>
<td>Poorly addressed</td>
<td>5.5</td>
</tr>
<tr>
<td>Carlbring et al. (2007)</td>
<td>Well-covered</td>
<td>Adequately addressed</td>
<td>Well-covered</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>Well-covered</td>
<td>Not reported</td>
<td>Adequately addressed</td>
<td>Poorly addressed</td>
<td>6.0</td>
</tr>
<tr>
<td>Furnark et al. (2009)</td>
<td>Well-covered</td>
<td>Adequately addressed</td>
<td>Well-covered</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>Well-covered</td>
<td>Not reported</td>
<td>Adequately addressed</td>
<td>Poorly addressed</td>
<td>6.0</td>
</tr>
<tr>
<td>Lovell et al. (2008)</td>
<td>Well-covered</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>Not reported</td>
<td>Poorly addressed</td>
<td>Adequately addressed</td>
<td>Poorly addressed</td>
<td>Well-covered</td>
<td>Adequately addressed</td>
<td>5.0</td>
</tr>
<tr>
<td>Richards et al. (2003)</td>
<td>Well-covered</td>
<td>Adequately addressed</td>
<td>Well-covered</td>
<td>Adequately addressed</td>
<td>Not reported</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>6.5</td>
</tr>
<tr>
<td>Salkovskis et al. (2006)</td>
<td>Well-covered</td>
<td>Adequately addressed</td>
<td>Well-covered</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>Well-covered</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>8.0</td>
</tr>
<tr>
<td>Schneider et al. (2005)</td>
<td>Well-covered</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>Poorly addressed</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>5.5</td>
</tr>
<tr>
<td>van Bockx et al. (2005)</td>
<td>Well-covered</td>
<td>Adequately addressed</td>
<td>Well-covered</td>
<td>Adequately addressed</td>
<td>Not reported</td>
<td>Well-covered</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>Adequately addressed</td>
<td>7.0</td>
</tr>
</tbody>
</table>

(i) The assignment of subjects to treatment groups is randomized.
(ii) An independent concealment of allocation procedure is used.
(iii) The treatment and control groups are similar at the start of the trial, with baseline scores described and differences assessed.
(iv) The only apparent difference between groups is the treatment under investigation (i.e. adequate statistical control or adjustment for confounding factors).
(v) Primary outcome measures are evidenced to be both valid and reliable and psychometric values are specified by the authors.
(vi) Levels of attrition are reported and equivalent for treatment versus control.
(vii) Intention-to-treat (ITT) analyses are reported and missing values are imputed.
(viii) A power calculation is reported and sufficient power is achieved.
(ix) The intervention is both sufficiently defined and delivered as planned (i.e. demonstrates good fidelity).
(x) The trial demonstrates external validity in terms of evaluating the intervention for an appropriate duration and within a clinically relevant setting.
suggest that the ‘large’ effects from media-recruited studies may not generalize to clinical practice settings. However, three of the six more clinically representative studies included some participants with severe symptoms of depression or anxiety. As GSH is a ‘low-intensity’ intervention intended for mild to moderate symptoms, the inclusion of individuals with severe symptoms may have undermined effectiveness within these studies. Regardless of the recruitment method, the findings indicate that the effectiveness of GSH at longer-term outcome is yet to be established.

Our finding that GSH interventions are less effective for patients recruited through primary care referrals compared to patients who self-select through media advertisements is consistent with previous reviews of the depression literature (Churchill et al. 2002; Gellatly et al. 2007) and anxiety and depression more generally (Westen & Morrison, 2001). Gellatly et al. (2007) noted that the evidence base for self-help treatments for depression, identified within previous NICE guidelines (2004), stems almost exclusively from self-selected rather than clinical samples. Similarly, within the updated NICE guideline for depression (2009), the bulk of evidence proposed to support the effectiveness of GSH in reducing depressive symptoms when compared with waitlist control is based primarily on five studies (which were included within the 2004 NICE guideline as referred to by Gellatly et al. 2007) that are predominantly based upon self-selected rather than clinical samples. Seven of the 13 included studies within the present review recruited some or all of their sample by media advertisements and self-selection. Such recruitment methods often rely on individuals’ motivation levels, which potentially correspond to a slightly different demographic from those participants who are recruited within primary care settings. Most of the methodologically stronger studies within the current review recruited research participants from clinical populations and generally demonstrated weak or non-significant effects of GSH upon anxiety or depression, particularly where outcomes were considered at follow-up rather than only immediately post-treatment. These findings highlight that the effectiveness of GSH within primary care settings as an effective treatment for anxiety and depressive disorders is not yet established and underlines the need for clinical recommendations to make reference to the potential differential impact of recruiting people by media advertisements versus clinical practice.

A further issue that contributes to the ambiguity of GSH effectiveness relates to the degree of treatment fidelity within the reviewed studies. With the exception of Lovell et al.’s (2008) study, which thoroughly addressed the issue of treatment fidelity, the remaining studies only partially addressed treatment fidelity in terms of sufficiently defining the intervention and reporting that it was delivered as planned. Of the 13 reviewed studies, only five explicitly mentioned that GSH therapists received GSH-specific training prior to applying GSH interventions. Furthermore, only six studies provided detail on whether therapists received supervision while guiding the intervention. Lack of detail regarding treatment fidelity, therapist training and therapist supervision reduces confidence in findings and generalizability of these studies, whether or not they endorse GSH as an effective intervention.

Strengths of review

We attempted to limit the potential for publication bias by corresponding with authors of all included review articles, and also authors of key relevant reviews, to obtain any unpublished findings. The potential for subjective bias in methodological analysis was also limited because we independently rated the methodological quality of included review studies, producing a high degree of inter-rater reliability.

Limitations of review

The current review was restricted to articles published in English, some electronic databases were not included within the search and a necessarily finite number of search terms were explored, all of which may have inadvertently excluded potentially relevant studies.

Comparing and synthesizing findings across a heterogeneous mix of mental health problems, amounts of guidance, outcome measures and follow-up periods was not straightforward and led to some inherent limitations. To minimize heterogeneity, the current review was confined to studies that met strict inclusion and exclusion criteria, such as limiting included studies to those with a therapist input of no less than 30 min and no more than 3 h. Although some purported GSH studies have involved therapist input for a greater or lesser duration, for the purposes of definition and guided by recent relevant literature (e.g. Mead et al. 2005; Gellatly et al. 2007), the range of 30 min to 3 h of therapist input was interpreted to be a proportionate amount of input representative of a GSH intervention. The review also excluded studies in which ‘guidance’ consisted simply of assessment or monitoring in order to assess conservatively the effectiveness of GSH. Although such definitions of GSH introduce an element of subjective bias, this delineation was necessary to afford a greater degree of specificity and transparency regarding the GSH interventions that were reviewed. It is acknowledged that, by attempting to increase specificity, the resultant pool
of reviewed studies was relatively small and the meta-
analysis was based on only a small number of studies.

Implications for research, clinical practice and policy

Currently, a wide variety of formats and duration of
therapist input are all defined as GSH, such that GSH
interventions are interchangeably, though perhaps not
systematically, defined within a whole host of varying
terminology (e.g. self-help, minimal contact inter-
vention and supervised self-help). The current review
attempted to define GSH as clearly as possible, as an
intervention: ‘involving access to self-help materials
in the treatment of mild to moderate anxiety or
depressive disorders, guided by the active support
(comprising more than reminders or monitoring) of a
professional or paraprofessional therapist for no less
than 30 minutes and no more than three hours in
total’. Greater consensus regarding the definition of
GSH and its distinction from non-GSH would facil-
itate future systematic evaluations of the effectiveness
of such interventions.

Given the apparent limited effectiveness of GSH at
follow-up, among higher quality studies and among
studies that recruited patients from clinical popu-
lations, it seems prudent to reserve judgement upon
GSH effectiveness within clinical settings until the
evidence base is substantiated by further high-quality
clinically based research trials that examine longer-
term effectiveness outcomes. This has implications for
guideline panels and service managers. The NICE
guidelines for depression (2009) currently recommend
individual GSH for mild depression despite fairly
varied outcomes among studies with wide variations
in terms of populations, recruitment and study qual-
ity. Indeed, this heterogeneity is acknowledged within
an appendix of those NICE guidelines, which con-
cedes that, across five studies indicating evidence of
GSH effectiveness, there is ‘serious inconsistency’
with heterogeneity greater than 50%. In addition, the
effectiveness referred to within these five studies
pertains to treatment end-point, not to follow-up.
Together, such heterogeneity and lack of follow-up,
highlighted within the current review as differentially
impacting upon GSH effectiveness outcomes, under-
lines the importance of considering such factors when
assessing the evidence base for the effectiveness of
GSH. It is essential for future GSH studies and sub-
sequent guidance to use more specific, consensual
definitions of GSH and to reflect more fully upon
issues of heterogeneity, recruitment and follow-up to
provide greater clarity regarding the effectiveness of
specific types of intervention for specific populations.

As outlined within the good-practice guidance of
self-help within the Improving Access to Psy-
chological Therapies (IAPT) services (Baguley et al.
2010): ‘further research is required looking at the
efficacy of self-help both across the range of disorders
and also the manner in which it might be delivered
(e.g. guided vs. unsupported).’ Although such low-
intensity interventions clearly need to offer patients
choice, many GSH studies could be more rigorous
in terms of documenting treatment fidelity and pro-
viding training/supervision for GSH therapists. The
introduction by IAPT of Psychological Wellbeing
Practitioners (PWPs), who receive training and super-
vision, points towards greater standardization. There
is a need for appropriate evaluation and dissemination
of clinical GSH services to facilitate understanding
of efficacy and predictors of outcomes within the
demands of clinical services. This would be aided by
further qualitative research to inform our under-
standing of the relevance, acceptability and key
components of GSH provision from the perspective of
patients. It is likely that certain types of GSH provided
by suitably trained and supervised therapists would
be effective for certain difficulties, but the evidence
base does not yet provide this level of certainty.

It has been documented that there are ‘currently
unrealistic assumptions about the proportion of
patients who can benefit from guided self-help’
and Seekles et al. (2009) state the case for more effec-
tiveness research within routine clinical practice in
order to evaluate not only whether certain self-help
interventions work but also whether they work in
clinical settings. The current review’s findings suggest
that GSH effectiveness outcomes are influenced by
study quality, recruitment settings and timing of out-
come, underlining the importance of methodological
rigour in future GSH effectiveness research. It seems
reasonable to expect that GSH can be effective in
certain formats for certain clients. Thus, GSH should
remain an integral component of stepped care, but in
the context of a research focus that is more defined,
agreed and scrutinized.

Lovell et al. (2008) indicate that more effective
targeting of GSH interventions is required, with re-
search into predictors or moderators of treatment ef-
effect, due to a current lack of understanding about who
benefits from GSH. Research is beginning to indicate
the impact of patient factors upon self-help more
generally (e.g. MacLeod et al. 2009). Similarly, Lucock
et al. (2008) and Williams & Martinez (2008) acknowl-
dge that future studies should explore the impact of
non-specific therapist factors upon self-help outcomes.
Although there is suggestion that monitoring by the
therapist is as effective as more structured guidance
(Gellaty et al. 2007), further research (particularly
with regard to anxiety disorders) exploring whether
monitoring is as beneficial as active guidance to patients will be necessary to ensure the provision of optimal levels of practitioner support within the low-intensity GSH interventions of the stepped care model. Greater understanding of the effective components of GSH and of the populations who genuinely benefit from such interventions is necessary to appropriately inform future evidence-based use of GSH within clinical practice.

Conclusions
This systematic review of the effectiveness of CBT-based GSH interventions for anxiety and depressive disorders demonstrates that the current evidence is inconclusive: GSH seems to be effective at post-treatment and for less clinically representative populations, but has limited effectiveness within routine clinical settings and in the longer term. Studies that have indicated greater effectiveness of CBT-based GSH have tended to be of poorer quality, have often neglected to provide follow-up data and have been primarily based upon media-recruited participants rather than clinical samples. To ensure that clinical practice is informed by appropriate clinical research findings and to elucidate whether GSH is effective for anxiety and depressive disorders, three aims for future research are suggested: (i) greater consensus regarding what constitutes GSH; (ii) more high-quality studies that evaluate the effectiveness of well-defined GSH within representative primary care samples; and (iii) more studies that report differences between treatment and control groups not only immediately following intervention but also crucially at longer-term follow-up intervals.

Acknowledgements
Grateful thanks are extended to the numerous published authors who replied to our request for relevant unpublished literature.

Declaration of Interest
None.

References


### Appendix 1. Systematic review of guided self-help (GSH) for anxiety and depressive disorders: excluded studies and reasons for exclusion

<table>
<thead>
<tr>
<th>Study</th>
<th>Reason(s) for exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andersson et al. (2006)</td>
<td>‘Minimal contact’ intervention also involved group exposure therapy</td>
</tr>
<tr>
<td>Beck et al. (1994)</td>
<td>Not a CBT-based GSH intervention (cognitive therapy versus relaxation training)</td>
</tr>
<tr>
<td>Berger et al. (2009)</td>
<td>No follow-up period</td>
</tr>
<tr>
<td>Bilich et al. (2008)</td>
<td>‘Minimal contact’ condition did not involve guidance; i.e. purely assessment and monitoring</td>
</tr>
<tr>
<td></td>
<td>‘Assisted self-help’ condition involved 4 h of assistance/guidance (i.e. &gt; 3 h)</td>
</tr>
<tr>
<td>Carlbring et al. (2001)</td>
<td>‘Minimal contact’ condition less than minimum criterion of 30 min of guidance</td>
</tr>
<tr>
<td>Carlbring et al. (2003)</td>
<td>‘Minimal contact’ condition did not involve guidance</td>
</tr>
<tr>
<td>Christensen et al. (2004)</td>
<td>No assessment of a diagnosis or cut-off score to establish caseness</td>
</tr>
<tr>
<td>Christensen et al. (2006)</td>
<td>Self-help conditions did not involve guidance</td>
</tr>
<tr>
<td>Clarke et al. (2005)</td>
<td>‘Minimal contact’ condition did not involve guidance</td>
</tr>
<tr>
<td>den Boer et al. (2007)</td>
<td>Intervention based on IPT in addition to CBT</td>
</tr>
<tr>
<td></td>
<td>Therapist contact greater than 3 h (~10 h)</td>
</tr>
<tr>
<td>Febbraro (2005)</td>
<td>No assessment of a diagnosis or cut-off score to establish caseness</td>
</tr>
<tr>
<td>Gould &amp; Clum (1995)</td>
<td>‘Minimal contact’ condition did not involve guidance</td>
</tr>
<tr>
<td>Hegerl et al. (2010)</td>
<td>No follow-up data reported</td>
</tr>
<tr>
<td>Jamison &amp; Scogin (1995)</td>
<td>‘Minimal contact’ condition did not involve guidance</td>
</tr>
<tr>
<td>Kupshik &amp; Fisher (1999)</td>
<td>Non-randomized design</td>
</tr>
<tr>
<td>Lovell et al. (2003)</td>
<td>Non-randomized design</td>
</tr>
<tr>
<td>Lovell et al. (2006)</td>
<td>Therapist input greater than criterion of maximum 3 h (&gt;5 h)</td>
</tr>
<tr>
<td>Lucock et al. (2008)</td>
<td>Non-randomized design</td>
</tr>
<tr>
<td>Proudfoot et al. (2004)</td>
<td>‘Minimal contact’ condition did not involve guidance</td>
</tr>
<tr>
<td>Reeves (2010)</td>
<td>Non-randomized design</td>
</tr>
<tr>
<td></td>
<td>Therapist input greater than criterion of maximum 3 h</td>
</tr>
<tr>
<td>Reeves &amp; Stace (2005)</td>
<td>Non-randomized design</td>
</tr>
<tr>
<td>Seekles et al. (2009)</td>
<td>Multi-component intervention, not solely GSH</td>
</tr>
<tr>
<td>Smit et al. (2006)</td>
<td>Not solely a CBT-based GSH intervention</td>
</tr>
<tr>
<td></td>
<td>Therapist input greater than criterion of maximum 3 h</td>
</tr>
<tr>
<td>Sorby et al. (1991)</td>
<td>‘Minimal contact’ condition less than criterion of minimum 30 min of guidance</td>
</tr>
<tr>
<td>Titov et al. (2008a)</td>
<td>No follow-up period</td>
</tr>
<tr>
<td>Titov et al. (2008b)</td>
<td>No follow-up period</td>
</tr>
<tr>
<td>Titov et al. (2008c)</td>
<td>No follow-up period</td>
</tr>
<tr>
<td>van Straten et al. (2008)</td>
<td>No assessment of a diagnosis or cut-off score to establish caseness</td>
</tr>
<tr>
<td></td>
<td>Not a CBT-based GSH intervention (problem-solving therapy)</td>
</tr>
</tbody>
</table>

CBT, Cognitive-behavioural therapy; IPT, interpersonal therapy.