Does Internet-based guided-self-help for depression cause harm? An individual participant data meta-analysis on deterioration rates and its moderators in randomized controlled trials

D. D. Ebert[1]*, L. Donkin[2,3], G. Andersson[4,5], G. Andrews[6], T. Berger[7], P. Carlbring[8], A. Rozenthal[8], I. Choi[9], J. A. C. Laferton[10], R. Johansson[4,5], A. Kleiboer[3], A. Lange[11], D. Lehr[12], J. A. Reins[12], B. Funk[12], J. Newby[13], S. Perini[14], H. Riper[3], J. Ruwaard[3], L. Sheeber[15], F. J. Snoek[16,17], N. Titov[18], B. Ünlü Ince[3], K. van Bastelaar[16], K. Vernmark[4,19], A. van Straten[3], L. Warmerdam[3], N. Salsman[20] and P. Cuijpers[3]

[1] *Friedrich Alexander University Erlangen-Nuremberg, Erlangen, Germany;* [2] *The Brain and Mind Research Institute, University of Sydney, NSW, Australia;* [3] *Department of Clinical, Neuro and Developmental Psychology, VU University Amsterdam, the Netherlands;* [4] *Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden;* [5] *Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden;* [6] *Clinical Research Unit for Anxiety and Depression, School of Psychiatry, University of New South Wales at St Vincent's Hospital, Darlinghurst, NSW, Australia;* [7] *Department of Clinical Psychology and Psychotherapy, University of Bern, Bern, Switzerland;* [8] *Department of Psychology, Stockholm University, Sweden;* [9] *Brain and Mind Centre, University of Sydney, Sydney, Australia;* [10] *Psychologische Hochschule Berlin;* [11] *Department of Clinical Psychology, University of Amsterdam, Amsterdam, The Netherlands;* [12] *Leuphana University Lüneburg, Germany;* [13] *Clinical Research Unit for Anxiety and Depression, School of Psychiatry, University of New South Wales at St Vincent's Hospital, Darlinghurst, NSW, Australia;* [14] *Centre for Emotional Health, Macquarie University, Australia;* [15] *Oregon Research Institute, Eugene, Oregon, USA;* [16] *Department of Medical Psychology, VU University Medical Center Amsterdam, The Netherlands;* [17] *Academic Medical Center/University of Amsterdam, The Netherlands;* [18] *eCentreClinic and MindSpot Clinic, Department of Psychology, Macquarie University, Australia;* [19] *Psykologpartners, Private Practice, Linköping, Sweden;* [20] *School of Psychology, Xavier University, Cincinnati, USA*

Background. Almost nothing is known about the potential negative effects of Internet-based psychological treatments for depression. This study aims at investigating deterioration and its moderators within randomized trials on Internet-based guided self-help for adult depression, using an individual patient data meta-analyses (IPDMA) approach.

Method. Studies were identified through systematic searches (PubMed, PsycINFO, EMBASE, Cochrane Library). Deterioration in participants was defined as a significant symptom increase according to the reliable change index (i.e. 7.68 points in the CES-D; 7.63 points in the BDI). Two-step IPDMA procedures, with a random-effects model were used to pool data.

Results. A total of 18 studies (21 comparisons, 2079 participants) contributed data to the analysis. The risk for a reliable deterioration from baseline to post-treatment was significantly lower in the intervention *v.* control conditions (3.36 *v.* 7.60; relative risk 0.47, 95% confidence interval 0.29–0.75). Education moderated effects on deterioration, with patients with low education displaying a higher risk for deterioration than patients with higher education. Deterioration rates for patients with low education did not differ statistically significantly between intervention and control groups. The benefit–risk ratio for patients with low education indicated that 9.38 patients achieve a treatment response for each patient experiencing a symptom deterioration.

Conclusions. Internet-based guided self-help is associated with a mean reduced risk for a symptom deterioration compared to controls. Treatment and symptom progress of patients with low education should be closely monitored, as some patients might face an increased risk for symptom deterioration. Future studies should examine predictors of deterioration in patients with low education.

* Address for correspondence: D. D. Ebert, Ph.D., Friedrich Alexander University Erlangen-Nuremberg, Erlangen, Germany.
(Email: david.ebert@fau.de)

Introduction

Major depressive disorder (MDD) is not only highly prevalent (Alonso *et al.* 2004; Waraich *et al.* 2004; Kessler *et al.* 2005; Wittchen *et al.* 2011; Rozental *et al.* 2014) but also associated with substantial impairment

(Ustün *et al.* 2004; Saarni *et al.* 2007) and economic costs (Berto *et al.* 2000; Greenberg & Birnbaum, 2005; Smit *et al.* 2006).

Psychological treatments have been shown to be effective in the treatment of depression (Cuijpers *et al.* 2008*a*, 2014). However, not all benefit from these treatments and many affected individuals remain untreated (Kohn *et al.* 2004; Wittchen *et al.* 2011).

Internet-based guided self-help interventions might be an acceptable (Cavanagh *et al.* 2011), effective (Johansson & Andersson, 2012; Richards & Richardson, 2012), and cost-effective (Hedman *et al.* 2012) treatment alternative, that could provide treatment to individuals not reached so far (Ebert *et al.* 2015*a*). While researchers have consistently demonstrated positive effects of Internet-based guided self-help for depression both, for adults (Richards & Richardson, 2012) and youths (Ebert *et al.* 2015*b*), little is known about potential negative effects of Internet-based psychological treatments for depression (Boettcher *et al.* 2014; Ebert *et al.* 2014*a*; Rozental *et al.* 2014). This is not unique for Internet treatments, as limited data are also available regarding negative effects for psychotherapy in general (Barlow, 2010; Emmelkamp *et al.* 2014).

While in pharmacological outcome research the standard is to always evaluate both risks and benefits of an intervention (Willan *et al.* 1997; Curtin & Schulz, 2011) psychotherapy outcome research has so far mostly focused on treatment benefits only (Lilienfeld, 2007; Dimidjian & Hollon, 2010).

Among different potential negative effects of psychotherapy one particularly unfavorable outcome is deterioration of symptoms as a consequence of treatment. Evidence from uncontrolled psychotherapy outcome studies indicates that a substantial number of patients experience a symptom deterioration while being in psychotherapy. The proportion of patients with symptom deterioration in these uncontrolled studies range from 3% to 14% (Smith & Glass, 1977; Mohr, 1995; Hansen *et al.* 2006; Lambert *et al.* 2006). This phenomenon of 'the deterioration effect' has been noted even in the early years of psychotherapy research (Bergin, 1966; Garfield *et al.* 1971).

With regard to Internet-based self-help treatments, it could be argued that such interventions may be associated with an even greater risk for symptom deterioration than face-to-face approaches. For example, for some individuals a self-help approach might not be intense enough (Kiluk *et al.* 2011). Further, individuals might be overstrained by trying to apply psychotherapeutic self-help strategies. Some therapeutic techniques could be inappropriately implemented by participants without direct guidance from a therapist. These problems could result in a further aggravation of hopelessness in severely affected individuals. It could also be argued

that in face-to-face treatments it is much easier to observe and react to early signs of deterioration than it is via the Internet. Another potential negative effect could be that self-help treatments could lead to a delayed help-seeking, which could result in a further deterioration of symptoms, if the initial low-intensity self-help treatment should be not sufficient.

Despite the fact that the topic of potential negative effects of both Internet-based treatments (Kiluk *et al.* 2011; Boettcher *et al.* 2014; Rozental *et al.* 2014, 2015; Bengtsson *et al.* 2015) and face-to-face psychotherapy (Lilienfeld, 2007; Barlow, 2010; Dimidjian & Hollon, 2010; Linden, 2013; Ladwig *et al.* 2014) have recently gained attention in the literature, empirical evidence on potential negative effects drawn from randomized controlled trials (RCTs) is still almost absent (Lilienfeld, 2007).

RCTs are the most reasonable approach to determine whether a treatment is differentially associated with a deterioration in functioning or an increase in symptomatology (Dimidjian & Hollon, 2010). If a randomized trial shows that participants in the active condition show greater deterioration in functioning than those in the non-treatment control condition, one can confidently conclude that the deterioration was a consequence of therapy (Lilienfeld, 2007). However, given that the number of people deteriorating during treatment is expected to be small, randomized trials are mostly underpowered to examine this research question adequately.

Moreover, RCTs evaluating psychological treatments seldom report the number of patients who deteriorated during treatment, thus it is not possible to investigate mean deterioration effects found in randomized trials and its moderators using traditional meta-analytical approaches. Consequently, to the best of our knowledge, there is no meta-analytical review on symptom deterioration and its moderators within RCTs evaluating Internet-based treatments or psychological interventions for depression in general. Given the increasing popularity of Internet-based treatments in healthcare systems worldwide (Andersson *et al.* 2014), there is a pressing need to evaluate potential deterioration effects of Internet-based treatments.

Individual participant data meta-analysis (IPDMA) can overcome some of the limitations of conventional meta-analysis at the study level (Clarke, 2005; Riley *et al.* 2010). By collecting and pooling the primary data of individual trials, analyses can be conducted, which have not been reported in the original studies. Furthermore, trials designed to detect overall treatment effects have limited power to detect treatment × subgroup interactions (Brookes *et al.* 2004). By combining the primary data of multiple trials using an IPDMA approach, it is possible to obtain a large sample size with sufficient power to examine effects in relevant
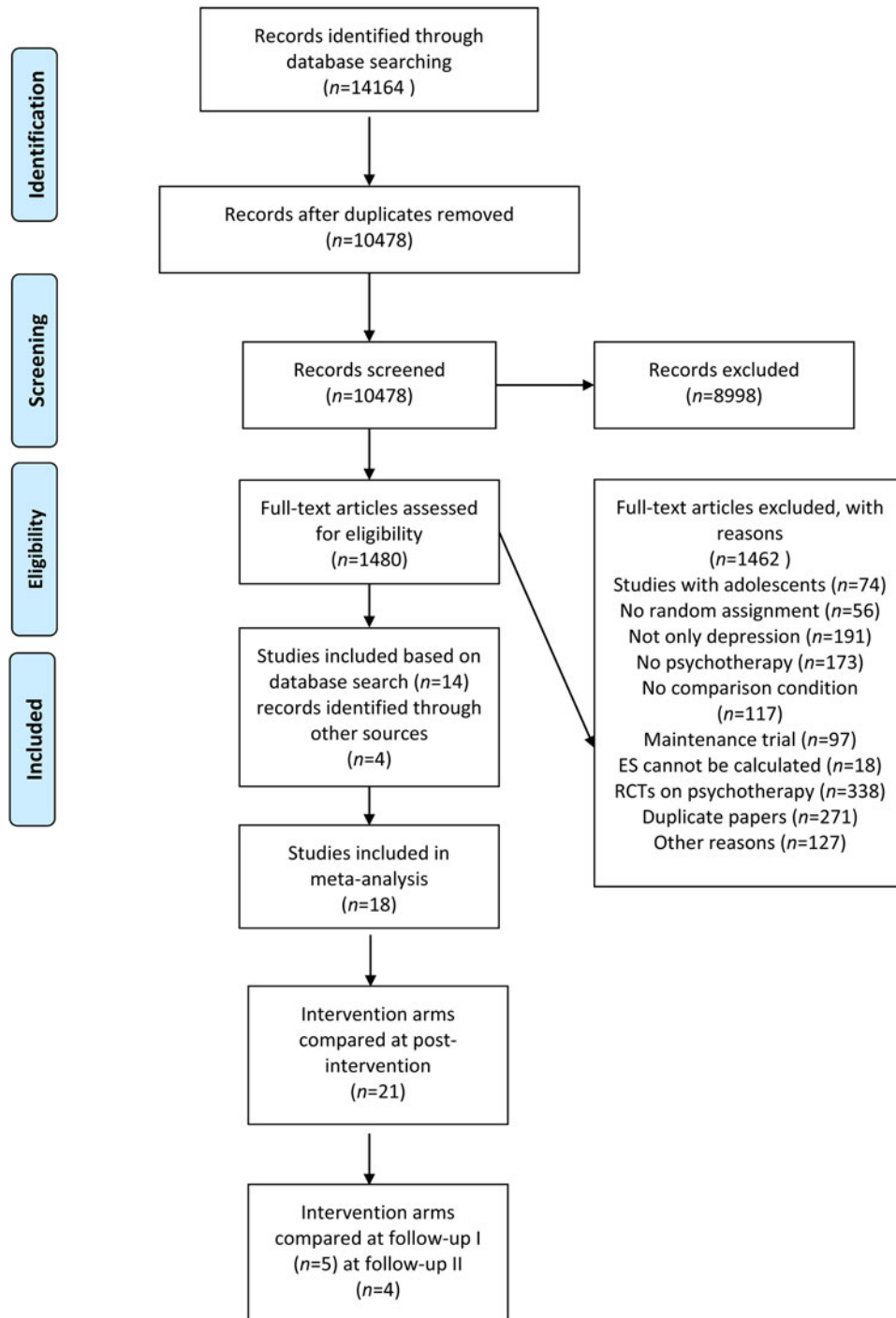
Fig. 1. Flowchart inclusion of studies.

subgroups and identify moderators of outcome (Cooper & Patall, 2009).

Hence, the present study aims to investigate deterioration rates and moderators of deterioration within randomized trials on Internet-based guided self-help interventions for adult depression, using IPDMA. We also evaluated deterioration rates in a number of subgroups of interest.

Method

Identification and selection of studies

In this study, we included randomized trials in which the effects of an Internet-based guided self-help treatment were compared with a control or comparison group (waiting list, care-as-usual, other) in adults (aged ⩾18 years) with depression (established by

diagnostic interview or elevated levels of depressive symptoms based on self-report measures). Studies were excluded if study participants were not currently in a depressive episode (e.g. if they were in remission), if the interventions were provided without guidance (i.e. without support from a therapist or other health-care professional) in order to increase internal validity and to reduce potential heterogeneity, if the interventions were delivered to the individual via a group format or were delivered at a location that required the individual to travel to use the programme (e.g. a clinic). Co-morbid general medical or psychiatric disorders were not used as a study exclusion criterion. No language restrictions were applied. Fig. 1 shows the selection process for included studies.

For the identification of potential studies for inclusion, we used a database of 1476 papers on the psychological treatment of depression that has been described in detail elsewhere (Cuijpers *et al.* 2008*b*). These searches covered papers published until January 2014 and in these searches we examined 14 164 abstracts in Pubmed (3638 abstracts), PsycInfo (2824), EMBASE (4682), and the Cochrane Central Register of Controlled Trials (3020). These abstracts were identified by combining terms indicative of psychological treatment and depression (both MeSH terms and text words). Further, the primary studies from 42 meta-analyses of psychological treatment for depression were checked to ensure that no published studies were missed. From the 14 164 abstracts (10 474 after removal of duplicates) 1476 full-text papers were retrieved for possible inclusion in the database.

Data collection, characteristics of included studies and participants

Corresponding authors were contacted for each of the identified papers and asked to provide raw data from their study. Of the 15 published studies identified from the database search, primary data was obtained from 14 (Andersson *et al.* 2005; van Straten *et al.* 2008; Warmerdam *et al.* 2008; Perini *et al.* 2009; Ruwaard *et al.* 2009; Vernmark *et al.* 2010; Titov *et al.* 2010; Berger *et al.* 2011; van Bastelaar *et al.* 2011; Choi *et al.* 2012; Johansson *et al.* 2012*a*, *b*; Sheeber *et al.* 2012; Ünlü Ince *et al.* 2013). Data for one study (Titov *et al.* 2011) could not be obtained, as the dataset was no longer available to the Titov research team. The study that was not included did not differ from the other studies in terms of design, participants, intervention, or quality. We also asked all authors whether they were aware of other recently completed RCTs that met our inclusion criteria, but were not yet published. Four more studies were identified by this method, and the authors were all willing to contribute their primary

data to this project (Carlbring *et al.* 2013; Newby *et al.* 2013; Ebert *et al.* 2014*c*; Kleiboer *et al.* 2015). This process resulted in a dataset with the primary data from 18 RCTs including 2079 cases. These 18 randomized controlled studies included 21 comparisons between an Internet-based guided self-help group *v.* control condition from baseline to post-test, five comparisons in addition from baseline to follow-up I (1–4 months, mean = 2.44, S.D. = 1.09, range 1–4, *n* = 737 participants) and four comparisons from baseline to follow-up 2 (⩾6 months, mean = 6.96, S.D. = 1.7, range 6–10, *n* = 594 participants). If a study had three conditions there would be two comparisons (i.e. the active treatment condition with each of the two control conditions). Only one study provided data for both follow-up time points (Ebert *et al.* 2014*b*). Characteristics of each included study are described in Table 1. Detailed information on sociodemographic and clinical characteristics of study participants can be found in Table 2.

Risk of bias assessment

The validity of included studies was assessed using four criteria of the 'Risk of Bias' assessment tool, developed by the Cochrane Collaboration (Higgins *et al.* 2011). This tool assesses possible sources of bias in randomized trials, including the adequate generation of allocation sequence; the concealment of allocation to conditions; the prevention of knowledge of the allocated intervention (masking of assessors); and dealing with incomplete outcome data (this was assessed as positive when intention-to-treat (ITT) analyses were conducted, meaning that all randomized participants were included in the analyses). Assessment of the quality was conducted independently by two assessors. Overall risk of bias was low. All studies reported an adequate sequence generation, and allocation to conditions by an independent (third) party. Sixteen studies reported blinding of outcome assessors or used only self-report outcomes, whereas five did not report blinding. All studies were coded as having handled missing data adequately, as ITT analyses were applied and missing data were imputed for all studies using multiple imputation. Sixteen studies met all four quality criteria, the remaining five studies met three of four criteria. Agreement between independent raters (P.C., L.D.) on the risk of bias was 95% across studies.

Missing data

Analyses were conducted according to the ITT principle. Missing data in the raw datasets were handled using multiple imputations (Schafer & Graham, 2002) with a Markov Chain Monte Carlo multivariate imputation algorithm (Missing data module in SPSS v. 20;

Table 1. *Selected characteristics of randomized controlled studies examining the effects of Internet-based psychotherapies for depression in adults*

Study	Recr	Depression	Inter-vention	N_{mod}	Time (weeks)	Guidance	N	Control group	N	Primary outcome	Duration of follow-up	Qual[a]	Publ	Country
1. Andersson *et al.* (2005)	Comm	MDD on CIDI-SF + MADRS-S 15–30	CBT	5	8	Feedback at the end of each module	62	Web-based discussion group	62	BDI-II	6 months	++++	1	SWE
2. Berger *et al.* (2011)	Comm	MDD on MINI + BDI-II >13	CBT	11	12	Weekly support via email by therapist	25	WL	26	BDI-II	6 months	++++	1	SWZ/GER
3. Carlbring *et al.* (2013)	Comm	MDD on MINI-DIS + PS BDI-II >13	ACT	7	13	Weekly contact by psychologist	40	WL	40	BDI-II	–	++++	0	SWE
4. Choi *et al.* (2012)	Comm	MDD on SCID-I PHQ-9 score >18	CBT	6	8	Weekly support via email or telephone by therapist	25	WL	30	BDI	N.A.	++ – +	1	AU
5. Ebert (2014b)	Comm	CES-D > 16	PST	5	5	Feedback at the end of each module	75	WL	75	CES-D	3 and 6 months	++++	0	GER
6. Johansson & Andersson (2012a)	Comm/Clin	MADRS-S 15–35 MDD confirmed by telephone interview	PD	9	10	Contact via online platform by therapist	46	Brief scheduled therapist support	46	BDI-II	10 months	++ – +	1	SWE
7. Johansson & Andersson (2012b)[a]	Comm/Clin	MADRS-S >14 MDD on SCID-I	CBT*	8–10	10	Contact via email by therapist	37	Moderated web-based discussion group	42	BDI-II	N/A	++++	1	SWE
							36							
8. Kleiboer (2015)	Comm	CES-D score of 16–39 HADS score of 8–14	PST	5	5	Contact via email by coach	106	WL	106	CES-D	N.A.	++++	0	NL
9. Newby *et al.* (2013)	Comm/Clin	MDD on MINI	CBT	6	10	Regular contact up to session 2, and response to user requests or decline in K10/PHQ-9 scores	25	WL	37	BDI-II	N.A.	++ – +	1	AU
10. Perini *et al.* (2009)	Comm	PHQ-9 score >4	CBT	6	6	Contact via email by therapist	27	WL	18	BDI-II	N.A.	++++	1	AU
11. Ruwaard *et al.* (2009)	Comm	BDI-IA score 10–29	CBT	8	11	Feedback on activities by therapist	36	WL	18	BDI-IA	N.A.	++++	1	NL
12. Sheeber *et al.* (2012)	Clin	Elevated self-reported levels of depression	CBT	8	8	Weekly contact via telephone	35	WL	35	BDI-II	N.A.	++++	1	USA
13. Titov *et al.* (2010)[a]	Comm	MDD on MINI	CBT	8	8	Weekly contact by therapist	38	WL	36	BDI-II	N.A.	++++	1	AU
			CBT	8	8		45							
14. Ünlü Ince (2013)	Comm	MDD on MINI CES-D >15	PST	5	5	Feedback on homework activities by coach	49	WL	47	CES-D	4 months	++++	0	NL

Table 1 (*cont.*)

Study	Recr	Depression	Inter-vention	N_{mod}	Time (weeks)	Guidance	N	Control group	N	Primary outcome	Duration of follow-up	Qual[a]	Publ	Country
15. Van Bastelaar et al. (2011)	Comm	MDD on CIDI CES-D >15	CBT	8	8	Feedback on homework activities by coach	125	WL	130	CES-D	1 month	++ − +	1	NL
16. Van Straten et al. (2008)	Comm	Self-defined depression or anxiety	PST	4	4	Feedback by coach	107	WL	106	CES-D	N.A.	++++	1	NL
17. Vernmark et al. (2010)	Comm	MDD on SCID-I-CV	CBT	7	7	Support via email by therapist	29	WL	29	BDI	N/A	++ − +	1	SW
18. Warmerdam et al. (2008)[a]	Comm	CES-D*	CBT	8	8	Weekly feedback by therapist	88	WL	87	CES-D	3 months	++++	1	NL
			PST	5	8		88							

Recr, Recruitment population; Comm, Community sample; Clin, Clinical Sample; Depression, confirmation of depression; CBT, cognitive behaviour therapy; ACT, acceptance and commitment therapy; PD, psychodynamic therapy; PST, problem-solving therapy; N_{mod}, Number of modules in the intervention; WL, waiting list control; BDI, Beck Depression Inventory; CES-D, Centre for Epidemiology Studies Depression Scale; Qual, risk of bias Score; Publ, publication of result (0 = unpublished, 1 = published); SWE, Sweden; SWZ, Switzerland; GER, Germany; NL, The Netherlands; AU, Australia; USA, United States of America.

[a] In this column a positive or negative sign is given for four quality criteria, respectively: allocation sequence; concealment of allocation to conditions; blinding of assessors; and intention-to-treat analyses.

Table 2. *Demographic and clinical characteristics*

	Intervention (N = 1123)			Control (N = 956)			All (N = 2079)		
	%	Mean	S.D.	%	Mean	S.D.	%	Mean	S.D.
Age, years		47.1	8.2		46.4	9.2		47.8	7.3
Female (%)	789 (70.57)			640 (66.04)			1419 (68.48)		
Married/partnership (%)	442 (49.44)			378 (47.61)			820 (48.58)		
Further education after high school (%)[a]	548 (59.05)			404 (54.16)			952 (56.87)		
BDI[b]									
Baseline		25.19	8.26		24.57	8.14		24.91	8.21
Post		13.24	8.98		21.01	10.09		16.78	10.26
Follow-up I[c]									
Follow-up II		12.23	8.25		14.52	8.81		13.39	8.6
CES-D[b]									
Baseline		28.86	8.4		28.46	8.42		28.67	8.41
Post		16.83	11.23		23.05	10.34		19.88	11.23
Follow-up I		17.80	8.75		22.72	8.34		20.05	8.91
Follow-up II		14.91	9.65		20.12	9.79		17.62	10.03
No current use of antidepressants[a]	27 (4.48)			11 (2.05)			38 (3.34)		
Comorbid anxiety disorder (%)[a]	308 (57.36)			288 (58.06)			596 (57.7)		

BDI, Beck Depression Inventory; CES-D, Centre for Epidemiological Studies Depression Scale.

[a] Percentages refer to those participants of studies who reported data.

[b] Medication: BDI and CES-D data refer to the imputed values.

[c] Studies that used the BDI as primary outcome did not assess follow-up I.

IBM Corp., USA) and 100 estimations per missing value. For the imputation of the primary outcome depression severity, we used all complete participant and study characteristics (study identifier, intervention group, baseline depression score, age, sex, recruitment population, confirmation of depression diagnosis method, intervention type, country of study, bias score – and post-intervention depression score when imputing follow-up). We did not impute baseline predictors.

Calculating deterioration rates

All studies used either the Centre for Epidemiological Studies – Depression Scale (CES-D; Radloff, 1977), or the Beck Depression Inventory (BDI; Beck *et al.* 1961) as outcome measures. Where multiple depression measures were present, the BDI was coded as the primary outcome measure given that it was the most frequently used outcome measure across studies. For both measures we calculated deterioration and response rates according to the widely used reliable change index (RCI; Jacobson & Truax, 1991). Participants whose scores from pre-treatment to post-treatment had RCIs below the cut point of −1.96 were considered to have experienced deterioration. A RCI of −1.96 is equivalent

to increases of depression of 7.68 points on the CES-D; and 7.63 points on the BDI.

Analyses

Effects of Internet-based treatments on deterioration rates were calculated using the standard two-step IPDMA approach (Riley *et al.* 2010). Thus, after calculating whether or not a participant deteriorated (yes/no) we calculated event rates for each study separately on the basis of the imputed data. Following this, pooled event rates across studies were calculated according to a random-effects model as implemented in the Comprehensive Meta-analysis software package version 2.2.021 (https://www.meta-analysis.com), accounting for clustering of both participants' within-study and between-study heterogeneity (Abo-Zaid *et al.* 2013). We proceeded by calculating the relative risks for each study, and pooled the results across the studies using a random-effects DerSimonian–Laird model (DerSimonian & Laird, 1986). For all analyses we chose a random-effects model, as we expected considerable heterogeneity among the studies. If there were significant differences between the groups with regard to deterioration, response, and remission rates, we also calculated the number needed to harm (NNH) and/or the number needed to treat (NNT)

and the associated 95% confidence intervals (CIs), compared to the control group. The NNH indicates the number of participants treated in the experimental condition for one extra person to demonstrate symptom deterioration as compared to the control group. We also calculated a benefit–risk ratio (Willan *et al.* 1997), by dividing the NNH for one extra symptom deterioration through the NNT to achieve one response [response was also defined using the reliable change criteria, such that participants with a reliable positive change (+1.96 on the RCI) were considered responders]. This procedure is usually used within drug treatment research (Curtin & Schulz, 2011), and quantifies the numbers of favourable outcomes achieved for each additional unfavorable outcome event incurred. Benefit–risk ratios were only calculated if there is a higher risk of deterioration in the intervention group as compared to the control group (Willan *et al.* 1997).

Sensitivity analyses

To test the robustness of our findings, we also conducted a sensitivity analysis applying an alternative criterion for deterioration. We defined the alternative criterion for deterioration such that individuals whose depression scores at baseline increased by $\geqslant 50\%$ at follow-up were categorized as having experienced deterioration. This criterion refers to a relative change instead of to an absolute change in symptoms.

Multiple treatments within one study

There were three studies in which two treatments were compared with a single control group (Warmerdam *et al.* 2008; Titov *et al.* 2010; Johansson *et al.* 2012*b*). In these cases, we treated each comparison as a separate study, and we avoided double counting of controls by randomly assigning half the control participants to each comparison.

Heterogeneity

As a test of homogeneity of effect sizes, we calculated the I^2 statistic as an indicator of heterogeneity in percentages (Ioannidis *et al.* 2007). A value of 0% indicates no observed heterogeneity, and larger values indicate increasing heterogeneity, with 25% as low, 50% as moderate, and 75% as high heterogeneity. We calculated 95% CIs around relative risks (RRs), using the non-central χ^2-based approach within the heterogi module for Stata (Orsini *et al.* 2013). We also calculated the Q statistic, but only report whether this was significant.

Publication bias

Publication bias was tested by inspecting the funnel plot and by Egger's test (Egger *et al.* 1997). We also applied Duval & Tweedie's trim-and-fill procedure (Duval & Tweedie, 2000), which yields an estimate of the effect size after the publication bias has been taken into account (Borenstein *et al.* 2009).

Subgroup analyses

We conducted a series of subgroup analyses. Pooling of the results was conducted according to the mixed-effects model. In this model, studies within subgroups are pooled with the random-effects model, while tests for significant differences between subgroups are conducted with the fixed-effects model. Subgroup analyses were only conducted for post-treatment data and not for follow-up data, as the sample sizes of follow-up datasets were not large enough to test for significant differences between subgroups. The following subgroups were investigated: *Participant characteristics*: sex (male/female); age group [adults (18–59 years), older adults ($\geqslant 60$ years)]; education [low (up to high school), medium to high (high school degree or further education after high school)]; co-morbid anxiety disorder (yes/no); depression severity at baseline [mild to moderate (BDI < 29); severe (BDI $\geqslant 29$)]; depression severity at baseline subgroup analyses was only calculated for participants of studies using the BDI, as the CES-D does not have an established cut-off score for depression severity. *Study characteristics*: MDD confirmed using an established diagnostic interview (yes/no); recruitment (community, clinical setting); risk of bias score [low (4); some risk (<4)]; type of control group (non-active/active). *Intervention characteristics*: theoretical model of the intervention (CBT, other); number of modules (4–5, 6–7, 8–11).

Ethical standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Results

Deterioration rates

Overall pooled reliable deterioration rates across measurements are summarised in Table 3. The risk for a reliable deterioration from baseline to post-treatment was significantly lower in the intervention *v.* control conditions (RR 0.47, 95% CI 0.29–0.75) and the NNT

Table 3. *Deterioration rates in Internet-based treatments of depression, sensitivity, and subgroup analyses*

		N_{co}	ER_{IG} (95% CI)	ER_{CG} (95% CI)	RR (95% CI)	Z	I^2 (95% CI)	NNT/NNH (95% CI)[a]	p[b]
Deterioration rates	Post-treatment	21	3.36 (2.5–5)	7.6 (5.59–9.70)	0.47 (0.29–0.75)	−3.17**	0 (0–50)	43.21 (25.83 to 132.10)	
(RCI)	Follow-up I	5	2.80 (2.5–5)	6.10 (2.9–12.3)	0.47 (0.20–1.42)	−1.66†	0 (0–79)	71.43 (21.28 to −52.63)	
	Follow-up II	4	6.0 (3.3–10.7)	5.30 (2.8–9.8)	1.17 (0.49–2.87)	0.35	0 (0–90)	−142.86 (30.3 to −21.28)	
Alternative deterioration criteria (50% increase)	Post-treatment	21	3.12 (2.09–4.62)	5.20 (3.64–7.37)	0.51 (0.29–0.86)	−2.50**	0 (0–52)	76.92 (38.46 to 1000)	
	Follow-up I	5	2.35 (0.75–7.14)	4.36 (1.96–9.42)	0.70 (0.28–1.75)	−0.76	0 (0–85)	111.11 (35.71 to −90.91)	
	Follow-up II	4	3.92 (1.93–7.83)	4.37 (2.28–8.19)	0.83 (0.30–2.27)	−0.37	0 (0–90)	100 (22.22 to −38.46)	
Outcome measures separately	Only BDI	14	3.37 (2.02–5.56)	8.51 (6.03–11.87)	0.39 (0.20–0.76)	−2.75**	0 (0–58)	27.03 (15.63 to 100)	
	Only CES-D	7	3.09 (1.52–6.16)	5.57 (3.12–9.74)	0.56 (0.29–1.11)	−1.66†	6.92 (0–76)	55.56 (23.81 to 200)	
Study characteristics									
Diagnosis	Confirmed MDD	11	3.63 (2.09–6.23)	8.07 (5.50–11.70)	0.42 (0.21–0.85)	−2.40*	0 (0–60)	23.26 (25.64 to 333.33)	0.17
	Depressive symptoms	10	3.13 (1.80–5.40)	6.18 (3.76–9.98)	0.51 (0.27–0.97)	−2.06*	3.78 (0–72)	55.56 (13.51 to 83.33)	
Target group	General populations	18	3.53 (2.36–5.24)	7.89 (6.05–10.24)	0.42 (0.25–0.70)	−3.36***	0 (0–54)	33.33 (20.83 to 83.33)	0.87
	Specific populations	3	3.66 (1.75–7.49)	4.69 (1.10–17.84)	0.84 (0.21–3.35)	−0.25	34.33 (0–79)	166.67 (14.93 to −18.18)	
Recruitment	Community recruitment	16	3.81 (2.61–5.53)	6.64 (4.69–9.33)	0.53 (0.32–0.89)	−2.21*	0 (0–57)	52.63 (28.57 to 1000)	0.27
	Clinical samples	5	2.28 (0.86–5.92)	9.46 (5.67–15.37)	0.26 (0.09 to −0.82)	−2.49*	0 (0–79)	16.67 (9.35 to 76.92)	
Risk of bias score	Some risk (<4)	5	2.42 (1.09–5.27)	6.62 (3.33–12.72)	0.50 (0.18–1.39)	−1.32	0 (0–79)	27.78 (11.90 to −83.33)	0.89
	Low (4)	16	3.92 (2.64–5.77)	7.70 (5.71–10.32)	0.47 (0.28–0.78)	−2.88**	0 (0–57)	40 (23.26 to 166.67)	
Type of control	Active	4	3.50 (1.52–7.83)	5.45 (2.75–10.53)	0.60 (0.20–1.76)	−0.94	0 (0–85)	40 (14.08 to −47.62)	0.21
	Non-active	17	3.57 (2.42–5.24)	7.52 (5.39–10.39)	0.45 (0.27–0.75)	−3.06**	0 (0–55)	38.46 (22.22 to 142.86)	
Intervention characteristics									
Intervention type	CBT	14	3.14 (1.96–5.00)	7.41 (5.07–10.71)	0.46 (0.25–0.87)	−2.39**	0 (0–58)	45.45 (22.73 to 2953.11)	0.91
	Other	7	3.35 (1.60–6.90)	7.05 (4.50–10.89)	0.49 (0.25–0.96)	−2.08*	0 (0–75)	37.04 (18.52 to −2821.37)	
No. of modules	4–5	8	4.23 (2.43–7.28)	7.33 (4.69–11.29)	0.58 (0.30–0·54)	−1.69	1.65 (0–71)	43.48 (19.23 to −166.67)	0.51
	6–7	3	1.61 (0.33–7.60)	7.59 (3.55–15.50)	0.18 (0.02–1.44)	−1.62	0[c]	22.73 (10 to −90.91)	
	8–11	10	2.85 (1.64–4.91)	6.90 (4.27–10.95)	0.41 (0.20–0.84)	−2.42*	0 (0–65)	45.45 (21.74 to −333.33)	
Patient characteristics									
Depression Severity	Mild-moderate	14	4.38 (2.53–7.48)	11.09 (7.43–16.23)	0.45 (0.23–0.94)	−2.12*	0 (0–60)	41.67 (17.24 to 90.91)	0.77
	Severe	14	5.37 (2.57–10.88)	9.20 (5.00–16.33)	0.36 (0.09–1.36)	−1.51	0 (0–79)	31.25 (11.11 to 38.46)	
Anxiety disorder	Yes	9	2.99 (1.50–5.87)	5.72 (3.47–9.28)	0.41 (0.16–1.01)	−1.87†	0 (0–54)	28.34 (15.31 to 190.53)	
	No	9	3.66 (1.70–7.69)	5.55 (2.84–10.56)	0.66 (0.22–1.97)	−0.75	0 (0–83)	238 (25.36 to −32.20)	0.52
Age group	Adults	21	3.51 (2.44–5.03)	7.97 (6.05–10.43)	0.46 (0.48–0.28)	−3.12**	0 (0–50)	45.45 (24.39 to 250)	0.84
	Old adults (⩾60)	19	7.56 (3.63–15.11)	10.67 (5.75–18.97)	0.39 (0.09–1.70)	−1.25	0 (0–85)	24.39 (9.17 to −38.46)	

Table 3 (*cont.*)

	N_{co}	ER_{IG} (95% CI)	ER_{CG} (95% CI)	RR (95% CI)	Z	I^2 (95% CI)	NNT/NNH (95% CI)[a]	p[b]
Sex								
Men	20	5.45 (3.24–9.03)	7.32 (4.75–11.12)	0.55 (0.26–1.17)	−1.56	0 (0–54)	40 (17.24 to −125)	0.68
Woman	21	3.51 (2.35–5.21)	8.27 (6.01–11.28)	0.45 (0.26–0.78)	−2.84**	0 (0–52)	52.63 (25.64 to −2989.24)	
Education								
Low (up to high school)	20	10.15 (5.34–18.45)	6.34 (2.66–14.45)	1.72 (0.53–5.61)	0.90	0 (0–79)	−30.30 (26.32 to −9.62)	0.03
Moderate/high (high school+)	21	3.66 (2.41–5.50)	8.07 (6.09–10.63)	0.39 (0.22–0.68)	−3.30***	0 (0–55)	31.25 (19.23 to 76.92)	

N_{co}, Number of comparisons; ER, event rate (number of patients with reliable deterioration); IG, intervention group; CG, control group; RR, relative risk for a deterioration; RCI, Reliable Change Index; BDI, Beck Depression Inventory; CES-D, Centre for Epidemiological Studies Depression Scale; MDD, major depressive disorder; CBT, cognitive behavioural therapy.

[a] A negative number in the column NNT/NNH indicates numbers needed to harm. A positive number indicates numbers needed to treat.

[b] The p value indicates whether differences between subgroups are significant.

[c] 95% CI around I^2 cannot be calculated as there must be at least three studies with patients experiencing a reliable deterioration to calculate 95% CIs.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, † $p < 0.1$.

to avoid one additional deterioration was 43.21 (95% CI 25.83–132.10). Heterogeneity was zero. The risk of a deterioration from baseline to follow-up I (1–4 months) appeared to demonstrate a trend towards being lower in the intervention groups compared to the control groups (RR 0.47, 95% CI 0.20–1.42), although the difference did not reach statistical significance ($p = 0.097$). There were no significant differences between the groups in the relative risk of a deterioration from baseline to follow-up II ($p = 0.72$). Heterogeneity was zero at both follow-ups.

Sensitivity and subgroup analyses are presented in Table 3. Applying the alternative deterioration criteria (50% symptom increase) resulted in slightly lower deterioration rates, but demonstrated a similar overall pattern.

Effects of Internet-based treatments compared to control conditions on deterioration rates at post-treatment were non-significant ($p > 0.05$) in 11 out of 25 subgroups tested and showed significantly better outcomes for those in Internet-based treatment in 14 out of 25 subgroups tested. RR was higher among those in Internet-based treatments than those in control conditions in only one subgroup (i.e. those with low education), although this difference was not significant (RR 1.72, NNH 30.3, 95% CI −26.32 to 9.62, $p = 0.37$).

Moderator of deterioration effects

Education level was also the only significant moderator of treatment effects on deterioration, such that there was significantly higher risk for deterioration for participants with lower levels of educational attainment compared to those with more education. All other differences between subgroups on deterioration rates were non-significant ($p > 0.10$).

Benefit–risk ratio

Participating in Internet-based treatments for depression was not associated with an increased risk for deterioration when compared to a control group. There was only one subgroup analysis (i.e. participants with low levels of education) in which the relative risk for a deterioration was higher in the intervention group compared to the controls, although this difference was non-significant. Nonetheless, analyses of response rates at post treatment for the subgroup of participants with low education showed that their responses to treatment were significantly higher in the intervention group compared to controls, with a relative risk of 1.91 and a NNT to achieve one additional response of 3.23 (Ebert *et al.* unpublished data). Dividing the NNH in order that one symptom deterioration occurs through the NNT to achieve one treatment response, results in a benefit–risk ratio of 9.38, indicating that 9.38 participants with low education

achieved a treatment response compared to the control group for each participant experiencing a deterioration in symptoms.

Publication bias

Inspection of the funnel plot and Egger's test indicated some possible publication biases. However, adjustment for publication bias using Duval & Tweedie's trim-and-fill procedure did not result in substantial changes. After adjustment for missing studies (five imputed studies) the RR for deterioration by post-test was 0.58 (95% CI 0.38–0.90), NNT were 43.21 (95% CI 25.83–132.10). Results at both follow-ups stayed the same.

Discussion

This IPDMA evaluated deterioration rates of Internet-based guided self-help interventions for depression compared to control conditions in randomized trials. In addition, deterioration rates were evaluated in subgroups of interest and potential moderating effects were examined.

Results showed that overall deterioration rates were low and the risk of deterioration was significantly lower for participants in Internet-based guided self-help conditions compared to controls. Education significantly moderated the risk for deterioration such that participants with lower educational attainment displayed a higher risk of deterioration compared to participants with more education. Nonetheless, a risk–benefit ratio analysis indicated that also in the subgroup of participants with low education the likelihood of benefits of positive response to Internet-based treatment clearly outweigh the possible risk for deterioration.

To the best of our knowledge, the present study is the first meta-analysis that evaluated deterioration effects and its moderators in RCTs evaluating a psychological treatment. Observed deterioration effects are in the lower range (3.6%) of those found in observational studies of face-to-face psychotherapy (3–14%) (Strupp *et al.* 1977; Mohr, 1995; Hansen *et al.* 2006; Lambert *et al.* 2006). In contrast to early indications of possible adverse effects of psychological treatments (Bergin, 1966; Garfield *et al.* 1971), we did not find a deterioration effect as a consequence of therapy. Instead results indicate that participating in Internet-based guided self-help programmes is associated with a lower risk of deterioration (RR 0.47) relative to controls. This effect held for the overall group and most subgroups. However, education level was identified as a significant moderator, with low educated participants at a greater risk for a deterioration than highly educated participants. This finding corresponds to results from some randomized trials that found that lower educational attainment was associated

with worse treatment outcomes compared to higher educated participants in Internet-based self-help interventions (Spek *et al.* 2008; Warmerdam *et al.* 2013). An explanation for such findings may be, that some patients with a lower educational level experience difficulties in terms of understanding the treatment modules, as most self-help manuals require a quite advanced reading comprehension. That may, in turn, decrease their self-efficacy and create feelings of hopelessness. Although all trials involved some form of guidance, this kind of support might not be sufficient for some individuals to overcome the barrier of low education. A more intensive treatment modality, as seeing a therapist face-to-face instead, could potentially help these patients understand the treatment rationale, and, thus, result in (hypothetically) less deterioration (Martinez *et al.* 2007). However, given that the topic of predictors of deterioration in psychotherapy has so far not been addressed in face-to-face psychotherapy, future studies should examine whether participants with a high risk for deterioration in Internet-interventions would be better suited for face-to-face psychotherapy. Another explanation may be that people with low education may also have other confounds (e.g. low income, poor physical health status, physical comorbidities, lower social support, less access to health services, etc.) which may either contribute to increased severity or lower ability to engage with the content/practice of skills from these programmes.

It should be noted, however, that deterioration rates of participants with low education (10%) were still in the range of those found in observational studies on face-to-face psychotherapy (3–14%, see above). Further, the benefit–risk ratio indicates that, in comparison to the control group, 9.38 patients with low education achieve a treatment response compared to the control group for each participant experiencing a deterioration in symptoms. It is also of note that with regard to response *v.* deterioration, a previously reported study using the same dataset did not find that education was a moderator of treatment response. Patients with low levels of education profited significantly and to almost the same extent (NNT = 3.23 for response) as patients with more education (NNT = 3.25 for response; Ebert *et al.* unpublished data). Thus, it is clear that most with low education experience response rather deterioration in Internet-based treatment. Therefore, low education alone should not be used to identify someone as high risk for deterioration and further research is needed to more specifically identify those who may be at high risk for deterioration.

When interpreting results from this study, several limitations need to be considered. First, the only negative effect evaluated was depression symptom deterioration. Other adverse effects may also occur and should be examined alongside RCTs in the context of

Internet-based guided self-help interventions in the future. For example, providing less intensive treatment than necessary through a self-help intervention might lead to lower treatment expectation in participants who fail to achieve a treatment response (Ebert *et al.* 2014*a*). Hence, although the present study did not find indications for harm of Internet-based self-help interventions, the study design does not allow to conclude an absence of harm. For a complete discussion on negative effects in Internet-based psychotherapy see Rozental *et al.* (2014) and for psychological treatments in general see Linden (2013). Future studies should examine other potential harmful effects of Internet-based treatments alongside RCTs. Second, given the limited number of studies that included a follow-up assessment, both for the intervention and the control condition, the analyses were underpowered to adequately conduct subgroup and moderator analyses at follow-up. Third, although the total number of participants was very high (2079), the low number of participants per subgroup did not allow for an examination of the association between the intervention and participant-characteristics within subgroups. For example, given the result for education as a moderator, future studies should investigate predictors of deterioration in the subgroup of participants with low levels of education in order to differentiate between those participants with low education with high chances for treatment success and participants with low education at high risk for failure. Moreover, all programmes were only examined in one randomized trial. Hence we were not able to investigate potential negative effects of specific programmes, which should be done in future studies. Fourth, the present study only investigated guided treatments for depression, hence we can not conclude anything about potential negative effects of unguided self-help treatments. Fifth, given the nature of IPDMA, the examined moderators of outcome were limited to those assessed in the original randomized trials. There might be other relevant moderators that have been not assessed.

The present study has relevant implications for both clinical practice and research. First, the differential results for moderators of effects on deterioration and treatment response indicate that the chances and risks for positive and negative change in psychological treatments might be two distinct constructs. Thus future studies on the differential effectiveness of psychological treatments should investigate moderators of both outcomes separately, instead of only evaluating the effects on mean change scores as is commonly done in psychotherapy research. Second, while many healthcare systems hesitate to implement Internet-based guided self-help approaches, the present study indicates, that such interventions are not associated with an increased risk for deterioration, but instead reduces the risk for a further aggravation of symptoms. Taken together with findings showing that such interventions have substantial positive effects on mean symptom improvement and on treatment response and remission (Johansson & Andersson, 2012; Ebert *et al.* 2015*b*) this further supports the need for dissemination of such treatments in routine mental health care. Nevertheless the moderator result for education indicates that a monitoring of participants with low education seems warranted, as they face an increased risk to deteriorate compared to participants with high education. Given that most participants with low education nevertheless achieve treatment response, and the mean effects on treatment response are comparable to those of participants with higher education, low education should not be used as an exclusion criteria in clinical practice. Instead, therapists should closely monitor the treatment and symptom progress in order to detect and react to early signs of a deterioration, e.g. by referral to more intensive treatment modalities. Given that the present study is the first study on deterioration rates in RCTs, one cannot conclude yet, whether these results are specific for Internet-based guided self-help intervention or whether such findings refer to psychological treatments for depression in general. Thus, future studies should evaluate deterioration rates and their moderators for face-to-face psychotherapy and should also compare overall and differential deterioration effects of internet-based and face-to-face psychotherapy.

The present study did not show any evidence for harm of Internet-based guided self-help interventions and indicates that such interventions reduce the risk for a symptom deterioration.

Acknowledgements

Declaration of Interest

None.

References

Abo-Zaid G, Guo B, Deeks JJ, Debray TPA, Steyerberg EW, Moons KGM, Riley RD (2013). Individual participant data meta-analyses should not ignore clustering. *Journal of Clinical Epidemiology* **66**, 865–873.

Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, de Girolamo G, Graaf R, Demyttenaere K,

Gasquet I, Haro JM, Katz SJ, Kessler RC, Kovess V, Lépine JP, Ormel J, Polidori G, Russo LJ, Vilagut G, Almansa J, Arbabzadeh-Bouchez S, Autonell J, Bernal M, Buist-Bouwman MA, Codony M, Domingo-Salvany A, Ferrer M, Joo SS, Martínez-Alonso M, Matschinger H, Mazzi F, Morgan Z, Morosini P, Palacín C, Romera B, Taub N, Vollebergh WAM (2004). Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatrica Scandinavica* (Suppl.), **420**, 21–27.

Andersson G, Bergström J, Holländare F, Carlbring P, Kaldo V, Ekselius L (2005). Internet-based self-help for depression: randomised controlled trial. *British Journal of Psychiatry* **187**, 456–461.

Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E (2014). Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. *World Psychiatry* **13**, 288–295.

Barlow DH (2010). Negative effects from psychological treatments: a perspective. *American Psychologist* **65**, 13–20.

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J (1961). An inventory for measuring depression. *Archives of GeneralPpsychiatry* **4**, 561–571.

Bengtsson J, Nordin S, Carlbring P (2015). Therapists' experiences of conducting cognitive behavioural therapy online vis-à-vis face-to-face. *Cognitive Behaviour Therapy* **44**, 470–479.

Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F (2011). Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. *Cognitive Behaviour Therapy* **40**, 251–266.

Bergin AE (1966). Some implications of psychotherapy research for therapeutic practice. *Journal of Abnormal Psychology* **71**, 235–246.

Berto P, D'Ilario D, Ruffo P, Di Virgilio R, Rizzo F (2000). Depression: cost-of-illness studies in the international literature, a review. *Journal of Mental Health Policy and Economics* **3**, 3–10.

Boettcher J, Rozental A, Andersson G, Carlbring P (2014). Side effects in internet-based interventions for social anxiety disorder. *Internet Interventions* **1**, 3–11.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2009). *Introduction to Meta-Analysis*. Wiley: Chichester, UK.

Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ (2004). Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *Journal of Clinical Epidemiology* **57**, 229–236.

Carlbring P, Hägglund M, Luthström A, Dahlin M, Kadowaki Å, Vernmark K, Andersson G (2013). Internet-based behavioral activation and acceptance-based treatment for depression: a randomized controlled trial. *Journal of Affective Disorders* **148**, 331–337.

Cavanagh K, Seccombe N, Lidbetter N (2011). The implementation of computerized cognitive behavioural therapies in a service user-led, third sector self-help clinic. *Behavioural and Cognitive Psychotherapy* **39**, 427–442.

Choi I, Zou J, Titov N, Dear BF, Li S, Johnston L, Andrews G, Hunt C (2012). Culturally attuned Internet treatment for depression amongst Chinese Australians: a randomised controlled trial. *Journal of Affective Disorders* **136**, 459–468.

Clarke MJ (2005). Individual patient data meta-analyses. *Best Practice and Research. Clinical Obstetrics and Gynaecology* **19**, 47–55.

Cooper H, Patall EA (2009). The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychological Methods* **14**, 165–176.

Cuijpers P, Karyotaki E, Weitz E, Andersson G, Hollon SD, van Straten A (2014). The effects of psychotherapies for major depression in adults on remission, recovery and improvement: a meta-analysis. *Journal of Affective Disorders* **159**, 118–126.

Cuijpers P, van Straten A, Andersson G, van Oppen P (2008*a*). Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology* **76**, 909–922.

Cuijpers P, van Straten A, Warmerdam L, Andersson G (2008*b*). Psychological treatment of depression: a meta-analytic database of randomized studies. *BMC Psychiatry* **8**, 36.

Curtin F, Schulz P (2011). Assessing the benefit: risk ratio of a drug–randomized and naturalistic evidence. *Dialogues in Clinical Neuroscience* **13**, 183–190.

DerSimonian R, Laird N (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177–188.

Dimidjian S, Hollon SD (2010). How would we know if psychotherapy were harmful? *American Psychologist* **1**, 21–33.

Duval S, Tweedie R (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* **56**, 455–463.

Ebert DD, Berking M, Heber E, Riper H, Laferton J, Cuijpers P, Lehr D (2015*a*). Restoring depleted resources: efficacy and mechanisms of change of an Internet-based unguided recovery training for better sleep and psychological detachment from work. *Health Psychology* **34** (Suppl.), 1240–1251.

Ebert DD, Lehr D, Baumeister H, Boß L, Riper H, Cuijpers P, Reins JA, Buntrock C, Berking M (2014*a*). GET.ON Mood Enhancer: efficacy of Internet-based guided self-help compared to psychoeducation for depression: an investigator-blinded randomised controlled trial. *Trials* **15**, 39.

Ebert DD, Lehr D, Boß L, Riper H, Cuijpers P, Andersson G, Thiart H, Heber E, Berking M (2014*c*). Efficacy of an internet-based problem-solving training for teachers: results of a randomized controlled trial. *Scandinavian Journal of Work, Environment and Health* **40**, 582–596.

Ebert DD, Zarski A-C, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, Riper H (2015*b*). Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. *PLoS ONE* **10**, e0119895.

Egger M, Davey Smith G, Schneider M, Minder C (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* (Clinical Research Edition) **315**, 629–634.

Emmelkamp PMG, David D, Beckers T, Muris P, Cuijpers P, Lutz W, Andersson G, Araya R, Banos Rivera RM, Barkham M, Berking M, Berger T, Botella C, Carlbring P, Colom F, Essau C, Hermans D, Hofmann SG, Knappe S, Ollendick TH, Raes F, Rief W, Riper H, Van Der Oord S, Vervliet B (2014). Advancing psychotherapy and evidence-based psychological interventions. *International Journal of Methods in Psychiatric Research* **23** (Suppl. 1), 58–91.

Garfield SL, Prager RA, Bergin AE (1971). Evaluating outcome in psychotherapy: a hardy perennial. *Journal of Consulting and Clinical Psychology* **37**, 320–322.

Greenberg PE, Birnbaum HG (2005). The economic burden of depression in the US: societal and patient perspectives. *Expert Opinion on Pharmacotherapy* **6**, 369–376.

Hansen NB, Lambert MJ, Forman EM (2006). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice* **9**, 329–343.

Hedman E, Ljótsson B, Lindefors N (2012). Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. *Expert Review of Pharmacoeconomics and Outcomes Research* **12**, 745–764.

Higgins JPT, Altman DG, Sterne JAC (eds) (2011). Chapter 8: Assessing risk of bias in included studies. In: *Cochrane Handbook for Systematic Reviews of Interventions, version 5.1.0 (updated March 2011)*, (updated March 2011) (ed. J. P. T. Higgins & S. Green). The Cochrane Collaboration, 2011 (www.cochrane-handbook.org). The Atrium, Southern Gate, Chichester.

Ioannidis JPA, Patsopoulos NA, Evangelou E (2007). Uncertainty in heterogeneity estimates in meta-analyses. *British Medical Journal (Clinical Research Edition)* **335**, 914–916.

Jacobson NS, Truax P (1991). Clinical significance: a statistical approach to denning meaningful change in psychotherapy research. *Psychology* **59**, 12–19.

Johansson R, Andersson G (2012). Internet-based psychological treatments for depression. *Expert Review of Neurotherapeutics* **12**, 861–869.

Johansson R, Ekbladh S, Hebert A, Lindström M, Möller S, Petitt E, Poysti S, Larsson MH, Rousseau A, Carlbring P, Cuijpers P, Andersson G (2012a). Psychodynamic guided self-help for adult depression through the internet: a randomised controlled trial. *PLoS ONE* **7**, e38021.

Johansson R, Sjöberg E, Sjögren M, Johnsson E, Carlbring P, Andersson T, Rousseau A, Andersson G (2012b). Tailored vs. standardized internet-based cognitive behavior therapy for depression and comorbid symptoms: a randomized controlled trial. *PLoS ONE* **7**, e36905.

Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry* **62**, 617–627.

Kiluk BD, Sugarman DE, Nich C, Gibbons CJ, Martino S, Rounsaville BJ, Carroll KM (2011). A methodological analysis of randomized clinical trials of computer-assisted therapies for psychiatric disorders: toward improved

standards for an emerging field. *American Journal of Psychiatry* **168**, 790–799.

Kleiboer A, Donker T, Seekles W, van Straten A, Riper H, Cuijpers P (2015). A randomized controlled trial on the role of support in internet-based problem solving therapy for depression and anxiety. *Behaviour Research and Therapy* **72**, 63–71.

Kohn R, Saxena S, Levav I, Saraceno B (2004). The treatment gap in mental health care. *Bulletin of the World Health Organization* **82**, 858–866.

Ladwig I, Rief W, Nestoriuc Y (2014). Welche Risiken und Nebenwirkungen hat Psychotherapie? – Entwicklung des Inventars zur Erfassung Negativer Effekte von Psychotherapie (INEP). *Verhaltenstherapie* **24**, 252–263.

Lambert MJ, Whipple JL, Hawkins EJ, Vermeersch DA, Nielsen SL, Smart DW (2006). Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science and Practice* **10**, 288–301.

Lilienfeld SO (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science* **2**, 53–70.

Linden M (2013). How to define, find and classify side effects in psychotherapy: from unwanted events to adverse treatment reactions. *Clinical Psychology and Psychotherapy* **20**, 286–296.

Martinez R, Whitfield G, Dafters R, Williams C (2007). Can people read self-help manuals for depression? a challenge for the stepped care model and book prescription schemes. *Behavioural and Cognitive Psychotherapy* **36**, 89–97.

Mohr DC (1995). Negative outcome in psychotherapy: a critical review. *Clinical Psychology: Science and Practice* **2**, 1–27.

Newby JM, Mackenzie A, Williams AD, McIntyre K, Watts S, Wong N, Andrews G (2013). Internet cognitive behavioural therapy for mixed anxiety and depression: a randomized controlled trial and evidence of effectiveness in primary care. *Psychological Medicine* **43**, 2635–2648.

Orsini N, Higgins J, Bottai M, Buchan I (2013). Heterogi: Stata module to quantify heterogeneity in a meta-analysis.

Perini S, Titov N, Andrews G (2009). Clinician-assisted Internet-based treatment is effective for depression: randomized controlled trial. *Australian and New Zealand Journal of Psychiatry* **43**, 571–578.

Radloff LS (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement* **1**, 385–401.

Richards D, Richardson T (2012). Computer-based psychological treatments for depression: a systematic review and meta-analysis. *Clinical Psychology Review* **32**, 329–342.

Riley RD, Lambert PC, Abo-Zaid G (2010). Meta-analysis of individual participant data: rationale, conduct, and reporting. *British Medical Journal (Clinical Research Edition)* **340**, c221.

Rozental A, Andersson G, Boettcher J, Ebert DD, Cuijpers P, Knaevelsrud C, Ljótsson B, Kaldo V, Titov N, Carlbring P (2014). Consensus statement on defining and measuring negative effects of Internet interventions. *Internet Interventions* **1**, 12–19.

Rozental A, Boettcher J, Andersson G, Schmidt B, Carlbring P (2015). Negative effects of internet interventions: a

qualitative content analysis of patients' experiences with treatments delivered online. *Cognitive Behaviour Therapy* **44**, 223–236.

Ruwaard J, Schrieken B, Schrijver M, Broeksteeg J, Dekker J, Vermeulen H, Lange A (2009). Standardized web-based cognitive behavioural therapy of mild to moderate depression: a randomized controlled trial with a long-term follow-up. *Cognitive Behaviour Therapy* **38**, 206–221.

Saarni SI, Suvisaari J, Sintonen H, Pirkola S, Koskinen S, Aromaa A, Lönnqvist J (2007). Impact of psychiatric disorders on health-related quality of life: general population survey. *British Journal of Psychiatry* **190**, 326–332.

Schafer JL, Graham JW (2002). Missing data : our view of the state of the art. *Psychological Methods* **7**, 147–177.

Sheeber LB, Seeley JR, Feil EG, Davis B, Sorensen E, Kosty DB, Lewinsohn PM (2012). Development and pilot evaluation of an Internet-facilitated cognitive-behavioral intervention for maternal depression. *Journal of Consulting and Clinical Psychology* **80**, 739–749.

Smit F, Cuijpers P, Oostenbrink J, Batelaan N, de Graaf R, Beekman A (2006). Costs of nine common mental disorders: implications for curative and preventive psychiatry. *Journal of Mental Health Policy and Economics* **9**, 193–200.

Smith ML, Glass GV (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist* **32**, 752–760.

Spek V, Nyklícek I, Cuijpers P, Pop V (2008). Predictors of outcome of group and internet-based cognitive behavior therapy. *Journal of Affective Disorders* **105**, 137–145.

Strupp HH, Hadley SW, Gomes-Schwartz B (1977). *Psychotherapy for Better or Worse: the Problem of Negative Effects.* Jason Aronson Inc. Publishers. New York.

Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K (2010). Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. *PLoS ONE* **5**, e10939.

Titov N, Dear BF, Schwencke G, Andrews G, Johnston L, Craske MG, McEvoy P (2011). Transdiagnostic internet treatment for anxiety and depression: a randomised controlled trial. *Behaviour Research and Therapy* **49**, 441–452.

Ünlü Ince B, Cuijpers P, van 't Hof E, van Ballegooijen W, Christensen H, Riper H (2013). Internet-based, culturally sensitive, problem-solving therapy for Turkish migrants

with depression: randomized controlled trial. *Journal of Medical Internet Research* **15**, e227.

Ustün TB, Ayuso-Mateos JL, Chatterji S, Mathers C, Murray CJL (2004). Global burden of depressive disorders in the year 2000. *British Journal of Psychiatry: the Journal of Mental Science* **184**, 386–392.

van Bastelaar KMP, Pouwer F, Cuijpers P, Riper H, Snoek FJ (2011). Web-based depression treatment for type 1 and type 2 diabetic patients: a randomized, controlled trial. *Diabetes Care* **34**, 320–325.

van Straten A, Cuijpers P, Smits N (2008). Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: randomized controlled trial. *Journal of Medical Internet Research* **10**, e7.

Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Oberg J, Carlbring P, Eriksson T, Andersson G (2010). Internet administered guided self-help versus individualized e-mail therapy: a randomized trial of two versions of CBT for major depression. *Behaviour Research and Therapy* **48**, 368–376.

Waraich P, Goldner EM, Somers JM, Hsu L (2004). Prevalence and incidence studies of mood disorders: a systematic review of the literature. *Canadian Journal of Psychiatry* **49**, 124–138.

Warmerdam L, Van Straten A, Twisk J, Cuijpers P (2013). Predicting outcome of Internet-based treatment for depressive symptoms. *Psychotherapy Research: Journal of the Society for Psychotherapy Research* **23**, 559–567.

Warmerdam L, van Straten A, Twisk J, Riper H, Cuijpers P (2008). Internet-based treatment for adults with depressive symptoms: randomized controlled trial. *Journal of Medical Internet Research* **10**, e44.

Willan AR, O'Brien BJ, Cook DJ (1997). Benefit-risk ratios in the assessment of the clinical evidence of a new therapy. *Controlled Clinical Trials* **18**, 121–130.

Wittchen H-U, Jacobi F, Rehm J, Gustavsson A, Svensson M, Jönsson B, Olesen J, Allgulander C, Alonso J, Faravelli C, Fratiglioni L, Jennum P, Lieb R, Maercker A, van Os J, Preisig M, Salvador-Carulla L, Simon R, Steinhausen H-C (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology* **21**, 655–679.