

## REVIEW

# Cost effectiveness of guided Internet-based interventions for depression in comparison with control conditions: An individual-participant data meta-analysis

Spyros Kolovos MSc<sup>1,2</sup>  | Johanna M. van Dongen PhD<sup>1</sup> | Heleen Riper PhD<sup>2</sup> |  
 Claudia Buntrock PhD<sup>2,3</sup> | Pim Cuijpers PhD<sup>2</sup>  | David D. Ebert PhD<sup>3</sup> |  
 Anna S. Geraedts PhD<sup>4</sup> | Robin M. Kenter PhD<sup>2</sup> | Stephanie Nobis PhD<sup>5</sup> |  
 Andrea Smith MSc<sup>6</sup> | Lianne Warmerdam PhD<sup>7</sup> | Jill A. Hayden PhD<sup>6</sup> |  
 Maurits W. van Tulder PhD<sup>1</sup> | Judith E. Bosmans PhD<sup>1</sup>

<sup>1</sup>Department of Health Sciences, Faculty of Earth and Life Sciences, Amsterdam Public Health Research Institute, VU University Amsterdam, Amsterdam, The Netherlands

<sup>2</sup>Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

<sup>3</sup>Department of Clinical Psychology and Psychotherapy, Institute of Psychology, University of Erlangen-Nürnberg, Nägelsbachstr, Erlangen, Germany

<sup>4</sup>ArboNed, Utrecht, The Netherlands

<sup>5</sup>Division of Online Health Training, Innovation Incubator, Leuphana University Lueneburg, Lueneburg, Germany

<sup>6</sup>Department of Community Health and Epidemiology, Faculty of Medicine, Dalhousie University, Halifax, Nova Scotia, Canada

<sup>7</sup>Stichting Benchmark GGZ (the Dutch Benchmark Foundation in Mental Health Care), Bilthoven, The Netherlands

## Correspondence

Spyros Kolovos, Department of Health Sciences and Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research Institute, VU University Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands.  
 Email: s.kolovos@vu.nl

## Funding information

FP7 Health, Grant/Award Number: 603098

**Background:** There is limited evidence on the cost effectiveness of Internet-based treatments for depression. The aim was to evaluate the cost effectiveness of guided Internet-based interventions for depression compared to controls.

**Methods:** Individual-participant data from five randomized controlled trials (RCT), including 1,426 participants, were combined. Cost-effectiveness analyses were conducted at 8 weeks, 6 months, and 12 months follow-up.

**Results:** The guided Internet-based interventions were more costly than the controls, but not statistically significant (12 months mean difference = €406, 95% CI: – 611 to 1,444). The mean differences in clinical effects were not statistically significant (12 months mean difference = 1.75, 95% CI: – .09 to 3.60 in Center for Epidemiologic Studies Depression Scale [CES-D] score, .06, 95% CI: – .02 to .13 in response rate, and .00, 95% CI: – .03 to .03 in quality-adjusted life-years [QALYs]). Cost-effectiveness acceptability curves indicated that high investments are needed to reach an acceptable probability that the intervention is cost effective compared to control for CES-D and response to treatment (e.g., at 12-month follow-up the probability of being cost effective was .95 at a ceiling ratio of 2,000 €/point of improvement in CES-D score). For QALYs, the intervention's probability of being cost effective compared to control was low at the commonly accepted willingness-to-pay threshold (e.g., at 12-month follow-up the probability was .29 and .31 at a ceiling ratio of 24,000 and 35,000 €/QALY, respectively).

**Conclusions:** Based on the present findings, guided Internet-based interventions for depression are not considered cost effective compared to controls. However, only a minority of RCTs investigating the clinical effectiveness of guided Internet-based interventions also assessed cost effectiveness and were included in this individual-participant data meta-analysis.

## KEYWORDS

cost effectiveness, cost utility, depression, individual-participant data meta-analysis, Internet-based intervention

## 1 | INTRODUCTION

Depression is a common mental health condition and poses a considerable disease burden in terms of years lived with disability (Vos et al., 2012). Major depressive disorder (MDD) has a 12-month prevalence of 5.1%, whereas the prevalence of subthreshold depression can be as high as 15% (Kessler & Bromet, 2013; Kessler et al., 2015; Shim, Baltrus, Ye, & Rust, 2011). Moreover, MDD is related to a substantial economic burden due to increased sickness absence, lost productivity while at work, and increased healthcare utilization (Lim, Sanderson, & Andrews, 2000). Thus, effective and cost-effective interventions for depression are needed to reduce the personal and societal burden caused by depression.

Guided Internet-based interventions constitute a promising, evidence-based form of treatment for individuals with depression (Andersson, Cuijpers, Carlbring, Riper, & Hedman, 2014; Andersson, Topooco, Havik, & Nordgreen, 2016). A small number of systematic reviews have evaluated the cost effectiveness of Internet-based interventions (Donker et al., 2015; Hedman, Ljotsson, & Lindefors, 2012; Tate, Finkelstein, Khavjou, & Gustafson, 2009). Even though the authors concluded that Internet-based interventions are a potentially useful clinical tool, they also concluded that there is a lack of evidence supporting their cost effectiveness. However, the RCTs included in these systematic reviews had different methodological approaches: they differed in the perspective used to estimate costs (e.g., societal or national healthcare provider), the valuation of cost categories, the outcome measures, and the analyses performed. As a result, review authors were unable to statistically pool the cost-effectiveness results in an aggregate data meta-analysis (Donker et al., 2015; Hedman et al., 2012; Tate et al., 2009).

To overcome these methodological issues, as well as the limited statistical power of cost-effectiveness analyses alongside RCTs, an individual-participant data meta-analysis (IPD-MA) can be used. IPD-MA is the gold standard of meta-analysis, since data are gathered for each individual participant from relevant studies and the outcomes can be defined consistently between studies (Riley, Lambert, & Abo-Zaid, 2010). The main objective of the current study was to use IPD-MA to assess, from a societal perspective, the cost effectiveness of guided Internet-based interventions compared to control conditions for adults with depressive symptoms.

## 2 | METHODS

### 2.1 | Search strategy and selection criteria

A systematic literature search was conducted in PubMed, Embase, PsycINFO, and the Cochrane Central Register of Controlled Trials from January 1, 2000 to January 1, 2017. We used key terms for depression, Internet-based treatments, and cost-effectiveness analyses (see Supporting Information). Furthermore, we asked experts and the authors of the selected studies if they were aware of other related studies. No language restriction was applied. There was no published protocol for this study. Additional information

on the methods of this study can be found in the Supporting Information.

Two reviewers (SK and JEB) independently examined the eligibility of the identified studies and reached consensus on which studies to include in the review. We included (a) randomized controlled trials, (b) comparing a guided Internet-based intervention, (c) to a control condition (e.g., care as usual, waiting list), (d) aimed at adults with depression or subclinical depression (based on a structured clinical interview or on elevated depressive symptomatology in a self-report measure), and (e) included measurements of depressive symptom severity, and healthcare utilization and productivity losses.

### 2.2 | Collection of IPD

We contacted the first authors of each study that met our inclusion criteria to ask for permission to access their primary dataset. We provided the authors with an overview of our study and a list of variables that we wished to acquire. We received data that were strictly anonymous making it impossible to track the identity of any of the participants.

### 2.3 | Risk of bias assessment

A risk of bias assessment of all included studies was conducted independently by two reviewers (SK and JEB) using the Cochrane Risk of Bias assessment tool (Higgins et al., 2011). The five risk of bias domains included in this tool are: random sequence generation (the randomization scheme was generated efficiently), allocation concealment (an independent party performed the allocation), blinding of outcome assessment (assessors were blind to treatment condition), incomplete outcome data (the method of handling incomplete outcome data was clearly described), and selective outcome reporting (all the outcomes reported in the study protocol or trial registry were presented in Section 3). "Blinding of participants and personnel" was omitted because concealment of treatment group is not possible due to the nature of the intervention.

To assess the quality of economic evaluation, we also assessed three domains related to risk of bias in economic evaluations (Evers, Hilgsmann, & Adarkwah, 2015): perspective (societal perspective was used), comparator (usual care was included as comparator), and cost measurements (all relevant cost categories were measured).

Each of the eight domains that were evaluated was scored as showing low or high risk of bias. If a study adhered to the operationalization described above, low risk of bias was considered present. Thus, each study could have a maximum score of 8 (i.e., a low risk of bias in all domains).

### 2.4 | Outcome measures

#### 2.4.1 | Depression

We differentiated between posttreatment (i.e., 8 weeks from baseline), short-term follow-up (i.e., 6 months from baseline), and long-term follow-up (12 months from baseline). Depressive symptomatology was assessed with the Center for Epidemiologic Studies Depression Scale

(CES-D) in all the included studies. Our primary clinical outcome of interest was change in depressive symptom severity and was calculated by subtracting the baseline CES-D score from the follow-up CES-D score. We also calculated the number of participants that positively responded to treatment at each follow-up. Response was defined as having a reduction in CES-D score from baseline to follow-up measurement of at least 50%.

### 2.4.2 | Quality of life

QoL was measured in all studies using the EuroQol questionnaire (EQ-5D-3L) (EuroQol-Group, 1990). Utility scores were calculated using the Dutch tariffs (Lamers, Stalmeier, McDonnell, Krabbe, & van Busschbach, 2005), since most of the studies were conducted in the Netherlands and to avoid differences due to different valuation tariffs. Utility scores are a preference-based measure of QoL, which are anchored at 0 (dead) and 1 (perfect health). Quality-adjusted life-years (QALYs) were calculated using the area-under-the-curve method, where the utility scores were multiplied by the amount of time an individual spent in a particular health state. Transitions between the health states were linearly interpolated.

### 2.4.3 | Costs

Costs were measured from the societal perspective, and all studies used adapted versions of the Trimbos and iMTA Questionnaire on Costs Associated with Psychiatric Illness (TiC-P) (Bouwman et al., 2013). TiC-P is a self-report questionnaire that measures healthcare utilization and productivity losses of patients with mental health problems. The questionnaires covered the full period between two measurement points (thus, the recall period was equal to the period between two measurements).

To prevent differences in costs between studies due to different prices, all resource utilization was valued using Dutch standard costs. All costs were expressed in Euros for the year 2014 using consumer price indices (Statistics-Nederlands, 2016). Details for healthcare costs, lost productivity, cost of the Internet-based interventions, and standard costs used to value them (Table S1) can be found in the Supporting Information.

## 2.5 | Statistical analysis

All analyses were performed using the pooled raw data (i.e., a one-stage approach) (Stewart & Parmar, 1993). The demographic and clinical characteristics of the patients were summarized using descriptive statistics. Separate analyses were conducted for each time point (8-week, 6-month, and 12-month follow-up) because we wanted to investigate whether our conclusions on the cost effectiveness of guided Internet-based treatment would change depending on the length of follow-up and, in addition, not all the included studies had the same measurement points.

The statistical analyses were carried out using the intention-to-treat principle. Missing data on costs and clinical effects for the various follow-up periods were imputed using multiple imputation by chained equations (MICE) as implemented in STATA 14 (Van Buuren, Boshuizen, & Knook, 1999). For this, we made the assumption that

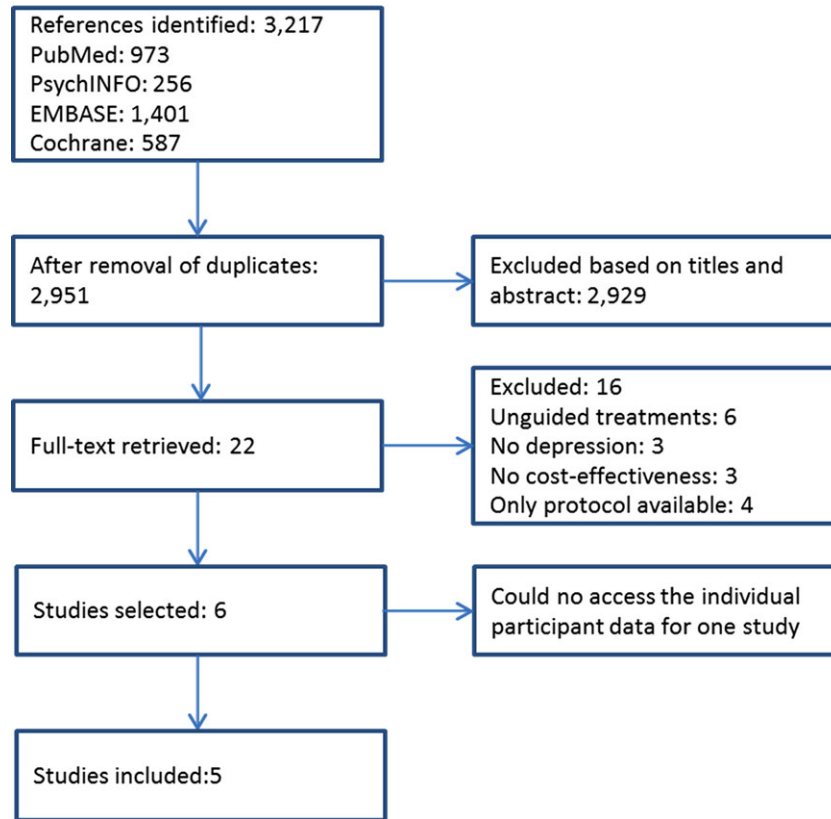
data were “missing at random” (Faria, Gomes, Epstein, & White, 2014). Imputations were performed separately for the intervention and the control group within each study. The imputation model contained variables related to “missingness” of data, variables differing at baseline between both groups and variables that predicted the outcome variables (including gender, age, depression severity and costs at baseline). Because the proportion of missing data was more than 10% and the majority of participants had missing data in all five EQ-5D domains, missing utility scores data were imputed at the index value level (Simons, Rivero-Arias, Yu, & Simon, 2015). We used predictive mean matching to account for the skewed distribution of costs (Vroomen et al., 2015). Twenty imputed datasets were needed in order for the loss of efficiency to be less than 5% (White, Royston, & Wood, 2011). The results of the 20 analyses were pooled according to Rubin's rules (2004). Rubin developed a set of rules to combine the estimates and standard errors from each imputed dataset into an overall estimate and standard error (Rubin, 2004).

Multilevel regression analyses were used to estimate the differences in costs and clinical effects between the intervention and control groups while accounting for the hierarchical structure of the data. A two-level structure was used where participants were nested within studies. To determine whether baseline variables were confounders, we used the “rule of thumb” of 10% change in the random coefficients between the model without covariates (crude model) and the model with covariates (adjusted model) (Maldonado & Greenland, 1993). As a result, we included baseline depression severity in the model to adjust for possible confounding for both costs and effects. We calculated the incremental cost-effectiveness ratios (ICERs) by dividing the difference in total costs between the intervention and the control group by the difference in effects. The 95% confidence intervals around the cost differences and the uncertainty surrounding the ICERs were estimated using bias-corrected bootstrapping with 5,000 replications (Efron, 1994).

To illustrate the statistical uncertainty surrounding ICERs, we plotted the bootstrapped cost and effect pairs on a cost-effectiveness plane (CE plane). In a CE plane, the incremental costs between the intervention and the control group are plotted on the y-axis and the incremental effects on the x-axis. We also estimated cost-effectiveness acceptability curves (CEA curves); they demonstrate the probability that the intervention is cost effective in comparison with the control group for a range of different values of willingness to pay (WTP), which was defined as the maximum amount of money society is willing to pay extra to gain one more unit of treatment effect (Fenwick, O'Brien, & Briggs, 2004). For this study, we used the commonly used UK WTP thresholds for QALYs (i.e., between 24,000 and 35,000 €/QALY) (NICE, 2013). These WTP thresholds are similar to the Dutch conventional WTP thresholds for depression (i.e., between 20,000 and 40,000 €/QALY) (Zwaap, Knies, van der Meijden, Staal, & van der Heiden, 2015).

## 2.6 | Sensitivity analyses

We conducted eight sensitivity analyses (at 12 months follow-up unless otherwise specified) to explore the robustness of our primary



**FIGURE 1** Study selection

analyses. We conducted the sensitivity analyses at 12-month follow-up because a longer follow-up measurement is more likely to capture long-term effects of the intervention (Ramsey et al., 2005). First, we repeated the analyses using the human capital approach (HCA) to value productivity losses (SA1). According to this approach, productivity losses cover the whole period during which the employee is absent from work. Second, we conducted the analyses from the national healthcare provider perspective in which only direct healthcare costs were included (SA2), because this is the recommended perspective in some countries, such as the United Kingdom (NICE, 2013). In the third and fourth sensitivity analyses, we decreased and increased the costs by 25% (SA3 and SA4), because the prices for healthcare utilization may be different in other countries as compared to the Netherlands.

In the fifth sensitivity analysis, we performed a complete case analysis to examine whether the results were similar to those of the imputed analyses (SA5). In addition, a per-protocol analysis including only participants that completed at least 75% of the online modules of the Internet-based treatment was performed (SA6). This operationalization of treatment completion has been used in a previous IPD-MA evaluating the adherence to Internet-based treatments, and was chosen based on the assumption that these participants were exposed to most of the core elements of the intervention (Karyotaki et al., 2015). We hypothesized that higher adherence to the intervention is associated with larger clinical effects. In the seventh sensitivity analysis, we created a more homogeneous sample by including studies that recruited participants from the general population, thus exclud-

ing participants from specialized mental healthcare or participants recruited at an occupational setting (SA7). This is the only sensitivity analysis performed at 6 months follow-up because at 12 months only one study recruited participants from the general population. Finally, we performed an analysis using the reliable change index as an outcome (SA8). The reliable change index is an indicator that the change in depressive symptoms from baseline to follow-up is larger than what would be expected due to random variation alone, and it was estimated based on the recommendations of Jacobson and Truax (1991). In this study, reliable change was defined as a reduction of at least eight points in CES-D between the baseline and the follow-up measurement.

### 3 | RESULTS

The search and study selection procedure is summarized in Figure 1. After removing duplicates, 2,951 papers were identified. We screened the titles and abstracts, and subsequently retrieved the full-text reports of 22 studies. Based on the full-text examination, we excluded 16 studies. We contacted the authors of the six eligible studies. The authors of five studies agreed to share the IPD from their trials with us (Buntrock et al., 2017; Geraedts et al., 2015; Kolovos et al., 2016; Nobis et al., 2013; Warmerdam, Smit, van Straten, Riper, & Cuijpers, 2010). The authors from one study were not able to share their data due to obstacles related to the informed consent that participants had

signed for the primary study, although they were willing to participate (Titov et al., 2015). Therefore, we eventually included five studies in this IPD-MA.

The characteristics of the five included studies are presented in Table 1. Three studies recruited participants from the general population, one study from specialized mental health care, and one study from an occupational setting. All studies used the CES-D to monitor the severity of depressive symptoms and the EQ-5D to measure QoL. Four studies included participants based on their CES-D scores and one study included participants with major depression based on DSM-IV criteria, which were evaluated by a structured clinical interview (Table 1). Finally, in one study participants had depression comorbid with diabetes (Nobis et al., 2013). The intervention and control condition of each study are reported in Table 1. Three of the studies were conducted in the Netherlands and two in Germany. We did not detect any inconsistencies in the data we received as compared to those reported in the primary studies (Buntrock et al., 2017; Geraedts et al., 2015; Kolovos et al., 2016; Nobis et al., 2013; Warmerdam et al., 2010).

The five studies included a total of 1,426 participants, 759 in the intervention groups and 667 in the control groups (Table S2 in Supporting Information). The mean age of the participants was 45 years (standard deviation [SD] = 12), 937 (66%) were female, 944 (66%) were employed, and 755 (53%) were married or living together with a partner. From the 1,426 participants, 100 (7%) attained low education, 454 (32%) middle education, and 871 (61%) higher education. At baseline, the mean CES-D score was 29 (SD = 9) and the mean utility score .65 (SD = .23) for the intervention group; for the control group, the mean CES-D score was 28 (SD = 9) and the mean utility score .64 (SD = .24).

Four of the studies were judged to have a low risk of bias in all eight domains (Table 1). One study showed low risk of bias in seven domains and a high risk of bias in one domain. This domain was the “comparator,” because this study included a waiting list group as a comparator (Table 1).

The unadjusted multiply imputed clinical and cost outcomes are presented in Table 2. Total societal costs were statistically nonsignificantly higher in the intervention as compared to the control at all three measurement points; at 8 weeks mean difference = €282, 95% CI: – 98 to 659, at 6 months mean difference = €378, 95% CI: – 343 to 1,040, and at 12 months mean difference = €504, 95% CI: – 1,149 to 1,892. Productivity losses were the largest contributor to total costs (Table 2).

The results of the cost-effectiveness analyses are presented in Table 3. The analyses at posttreatment (i.e., 8 weeks) included 763 participants from three studies (Geraedts et al., 2015; Kolovos et al., 2016; Warmerdam et al., 2010). The ICER for improvement in depressive symptoms was 224, meaning that one additional point of improvement in CES-D score in the intervention group was on average associated with €224 higher costs as compared to the control group (Table 3). The ICER for response was 3,903, meaning that for one additional participant responding to treatment in the intervention group compared to the control group an investment of on average €3,903 is needed (Table 3). For QALYs, the ICER was 81,555, meaning that one QALY gained in the intervention was associated with an extra cost of on average €81,555 in comparison with the control (Table 3). The results from

**TABLE 1** Characteristics of the included studies

Author	Recruitment	Primary Diagnosis	Intervention (n)	Control (n)	Depression Measure	Costs Measure	Follow-Up <sup>b</sup>	Risk of Bias	Country
Buntrock (2015)	General population <sup>a</sup>	Depressive symptoms $\geq 16$ CES-D and no MDD based on DSM-IV	GBT+CAU (202)	CAU+PE (204)	CES-D	TiC-P	26, 52	8/8	GER
Geraedts (2016)	Occupational setting	Depressive symptoms $\geq 16$ CES-D	GBT+PST (116)	CAU (115)	CES-D	TiC-P	8, 26, 52	8/8	NL
Kolovos (2016)	Specialized mental health care	MDD diagnosis based on DSM-IV	PST+CAU (136)	Self-help book+CAU (133)	CES-D	TiC-P	8, 26, 52	8/8	NL
Nobis (2016)	General population	Depressive symptoms $\geq 23$ CES-D and diabetes	GBT+CAU (129)	CAU+PE (128)	CES-D	TiC-P	26	8/8	GER
Warmerdam (2010)	General population	Depressive symptoms $\geq 16$ CES-D	PST (88), CBT (88)	WL (87)	CES-D	TiC-P	5, 8	7/8	NL

CAU = care as usual; CBT = cognitive behavioral therapy; CES-D = Center for Epidemiological Studies for Depression Scale; GER = Germany; MDD = major depressive disorder; NL = Netherlands; PE = psychoeducation; PST = problem solving therapy; TiC-P = Trimbos and IMTA questionnaire on costs associated with psychiatric illness; WL = waiting list.

<sup>a</sup>Participants with subclinical depression.

<sup>b</sup>In weeks from baseline.

**TABLE 2** Multiply imputed unadjusted mean costs (€, 2014) at 8 weeks, 6 months, and 12 months follow-up

Outcome	Intervention Group	Control Group	Mean Difference (95% CI)
8 weeks (n = 763)			
Clinical effects			
Improvement in CES-D	9.8	8.3	1.5 (– .4 to 3.4)
Response rate	.22	.19	.04 (– .02 to .10)
QALY	.10	.10	0 (– .01 to .01)
Cost categories			
Intervention	268	0	268 (N/A)
Mental health	152	143	9 (– 33 to 52)
Primary care	76	73	3 (– 19 to 24)
Secondary care	108	120	–12 (– 82 to 56)
Medication	1	2	–1 (– .5 to 1.5)
Complementary therapy	11	10	1 (– 5 to 8)
Domestic help	391	380	11 (– 103 to 123)
Productivity losses	1,061	1,058	3 (– 305 to 309)
Total	2,068	1,786	282 (– 98 to 659)
6 months (n = 1,163)			
Clinical effects			
Improvement in CES-D	9.24	7	2.24 (.8–3.7)
Response rate	.20	.18	.02 (– .02 to .07)
QALY	.36	.35	.01 (– .01 to .02)
Cost categories			
Intervention	243	5	238 (233–242)
Mental health	268	234	34 (– 115 to 46)
Primary care	164	129	35 (0–70)
Secondary care	482	334	148 (– 187 to 485)
Medication	5	5	0 (– 2 to 2)
Complementary therapy	18	22	–4 (– 15 to 6)
Domestic help	907	750	157 (– 65 to 379)
Productivity losses	2,350	2,580	–230 (– 758 to 161)
Total	4,437	4,059	378 (– 343 to 1,040)
12 months (n = 906)			
Clinical effects			
Improvement in CES-D	11.7	9.6	2.1 (.1–4)
Response rate	.27	.24	.3 (– .03 to .08)
QALY	.76	.76	.00 (– .3 to .3)
Cost categories			
Intervention	238	4	234 (230–239)
Mental health	563	646	–83 (– 272 to 134)
Primary care	270	239	31 (– 47 to 103)
Secondary care	731	429	302 (– 154 to 741)
Medication	10	11	–1 (– 6 to 5)
Complementary therapy	42	51	–9 (– 35 to 18)
Domestic help	2,059	1,712	347 (– 168 to 802)
Productivity losses	4,335	4,652	–317 (– 1,182 to 692)
Total	8,248	7,744	504 (– 1,149 to 1,892)

**TABLE 3** Differences in mean adjusted costs (€, 2014) and effects (95% confidence intervals), incremental cost-effectiveness ratios, and distribution of incremental cost-effect pairs on the cost-effectiveness planes

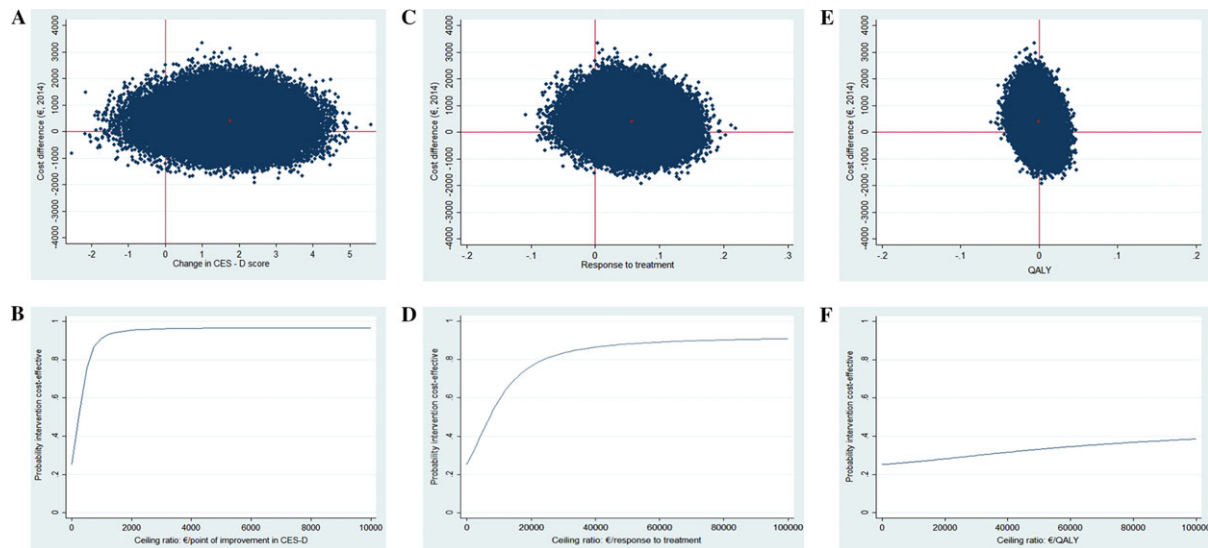
Analysis Outcome	$\Delta C$ (95% CI) €	$\Delta E$ (95% CI) Points	ICER €/Point	Distribution on CE Plane <sup>b</sup> (%)			
				NE	SE	SW	NW
Main analysis							
8 weeks ( <i>n</i> = 763)							
CES-D	400 (92; 705)	1.79 (-.12; 3.70)	224	96	1	0	3
Response	400 (92; 705)	.10 (.03; .17)	3,903	98	1	0	1
QALYs	400 (92; 705)	.01 (.00; .01)	81,155	98	1	0	1
6 months ( <i>n</i> = 1,163)							
CES-D	211 (- 355; 787)	2.17 (.70; 3.64)	97	74	26	0	0
Response	211 (- 355; 787)	.07 (.01; .13)	3,158	73	26	0	1
QALYs	211 (- 355; 787)	.01 (- .00; .02)	32,706	65	24	2	9
12 months ( <i>n</i> = 906)							
CES-D	406 (- 611; 1,444)	1.75 (-.09; 3.60)	232	72	25	1	3
Response	406 (- 611; 1,444)	.06 (-.02; .13)	7,079	68	24	2	6
QALYs	406 (- 611; 1,444)	-.00 (-.03; .03)	-1,122,646	33	16	9	42
Sensitivity analysis <sup>a</sup>							
SA1: Human capital approach							
CES-D	1,328 (- 32; 2,716)	1.75 (-.09; 3.60)	758	90	6	0	3
Response	1,328 (- 32; 2,716)	.06 (-.02; .13)	23,140	87	6	0	7
QALYs	1,328 (- 32; 2,716)	-.00 (-.03; .03)	-23,210,980	45	5	2	49
SA2: National healthcare provider							
CES-D	442 (117; 824)	1.76 (-.09; 3.60)	251	95	2	0	3
Response	442 (117; 824)	.06 (-.02; .13)	7,698	91	2	0	7
QALYs	442 (117; 824)	-.00 (-.03; .03)	-1,220,857	47	1	1	51
SA3: Increased costs by 25%							
CES-D	508 (- 764; 1,808)	1.75 (-.09; 3.60)	289	72	25	1	3
Response	508 (- 764; 1,808)	.06 (-.02; .13)	8,848	69	24	2	6
QALYs	508 (- 764; 1,808)	-.00 (-.03; .03)	-1,403,308	33	16	9	42
SA4: Decreased costs by 25%							
CES-D	305 (- 458; 1,082)	1.75 (-.09; 3.60)	174	72	25	1	3
Response	305 (- 458; 1,082)	.06 (-.02; .13)	5,039	69	24	2	6
QALYs	305 (- 458; 1,082)	-.00 (-.03; .03)	-841,985	33	16	9	42
SA5: Complete case ( <i>n</i> = 482)							
CES-D	-62 (- 1,304; 1,202)	2.7 (.92; 4.62)	-23	47	53	0	0
Response	-62 (- 1,304; 1,202)	.12 (.03; .20)	-539	47	53	0	0
QALYs	-62 (- 1,304; 1,202)	.01 (-.01; .03)	-8,365	31	42	11	17
SA6: Per protocol ( <i>n</i> = 681)							
CES-D	279 (- 935; 1,524)	2.36 (.41; 4.62)	118	66	34	0	1
Response	279 (- 935; 1,524)	.07 (-.02; .16)	4,070	62	32	2	5
QALYs	279 (- 935; 1,524)	.01 (-.02; .03)	39,102	44	28	6	22
SA7: Recruited from general population <sup>c</sup> ( <i>n</i> = 663)							
CES-D	59 (- 673; 750)	3.17 (1.84; 4.51)	18	55	45	0	0
Response	59 (- 673; 750)	.12 (.05; .18)	504	55	45	0	0
QALYs	59 (- 673; 750)	.01 (-.01; .02)	15,437	39	34	11	16
SA8: Reliable change index	406 (- 611; 1,444)	.07 (-.01; .14)	5,971	71	26	1	3

Notes: CES-D = Center for Epidemiologic Studies Depression Scale (range:0-6-); QALYs = Quality-Adjusted Life-years (range: 0-1);  $\Delta C$  = Mean difference in costs;  $\Delta E$  = Mean difference in effects.

<sup>a</sup>Sensitivity analyses conducted at 12-month follow-up. *N* = 906 unless otherwise specified.

<sup>b</sup>The northeast (NE) quadrant of the CE plane indicates that the Internet-based intervention is more effective and more expensive than the control condition; The southeast (SE) quadrant of the CE plane indicates that the Internet-based intervention is more effective and less expensive than the control condition; The southwest (SW) quadrant of the CE plane indicates that the Internet-based intervention is less effective and more expensive than the control condition; The northwest (NW) quadrant indicates that the Internet-based intervention is less effective and more expensive than the control condition.

<sup>c</sup>This sensitivity analysis was conducted at 6 months follow-up because at 12 months only one study was conducted in primary care



**FIGURE 2** Cost-effectiveness plane at 12 months of Internet-based treatment versus control condition for (a) improvement in CES-D score, (c) response to treatment, and (e) QALYs. Cost-effectiveness acceptability curve for (b) improvement in CES-D score, (d) response to treatment, and (f) QALYs at 12 months

the CE planes and CEA curves are reported in detail the Supporting Information.

The analyses at 6-month follow-up included 1,163 participants from four studies (Table 3) (Buntrock et al., 2017; Geraedts et al., 2015; Kolovos et al., 2016; Nobis et al., 2013). The ICER for improvement in depressive symptoms was 97 (i.e., €97 per point improvement extra in CES-D score for intervention vs. control). The ICER for response was 3,158 (i.e., €3,158 per additional participant that responded to treatment for intervention vs. control) (Table 3). For QALYs, the ICER was 32,706 (i.e., €32,706 per QALY gained for intervention vs. control) (Table 3).

The analyses at 12-month follow-up included 906 participants from three studies (Table 3) (Buntrock et al., 2017; Geraedts et al., 2015; Kolovos et al., 2016). The ICER for depressive symptoms was 232 (i.e., €232 per point improvement extra in CES-D score for intervention vs. control). The ICER for response was 7,079 (i.e., €7,079 per additional participant that responded to treatment for intervention vs. control). The ICER for QALYs was – 1,122,646, meaning that one QALY lost in the intervention group was associated with an extra cost of on average €1,122,646 in comparison with the control (the ICER was large and negative because the mean difference in QALYs between the two groups was very small favoring the control group). The CE planes and CEA curves for each of the three outcomes are presented in Figure 2.

The cost-effectiveness results from the sensitivity analyses SA1, SA2, SA3, SA4, and SA8 were similar to those in the main analysis (Table 3). When we included only patients with complete data (SA5;  $n = 482$ ), total societal costs were lower (mean difference – 62; 95% CI – 1,304 to 1,202) and the clinical effects larger in the intervention group as compared to the main analysis. In the per-protocol analyses (SA6;  $n = 681$ ), the mean difference in total costs was smaller and in favor of the intervention as compared to the main analysis. The mean differences in effects were also larger for CES-D improvement and response as compared to the main analysis, again favoring the intervention. However, the probability that the intervention was cost

effective compared to control for both analyses (SA5 and SA6) did not change substantially, as compared to the main analysis.

Finally, in the analyses including studies that recruited participants from the general population (SA7;  $n = 663$ ), the probability that the intervention was cost effective compared to control was higher in these set of studies as compared to the respective probability from the main analyses. For instance, for QALYs the probability was .44 at a ceiling ratio of 0 €/QALY, .53 at a ceiling ratio of 24,000 €/QALY, and .56 at a ceiling ratio of 35,000 €/QALY.

## 4 | DISCUSSION

This IPD-MA evaluated the cost effectiveness of guided Internet-based interventions for depression compared to control conditions including a total of five studies, with measurements at 8 weeks (three studies,  $n = 763$ ), 6 months (four studies,  $n = 1,163$ ), and 12 months (three studies,  $n = 906$ ) of follow-up.

Based on the results, guided Internet-based interventions cannot be considered cost effective compared to control conditions for depressive symptoms, treatment response and QALYs at all three follow-up measurements. We identified better cost and effects estimates at the 6 months follow-up compared to 8 weeks and 12 months follow-ups. Nevertheless, as indicated at all three measurement points decision makers need to make considerable investments to reach an acceptable probability of cost effectiveness. For instance, at the commonly used UK WTP thresholds for QALYs (i.e., between 24,000 and 35,000 €/QALY) the probability of the intervention being cost effective compared to the control was relatively low (e.g., at 12-month follow-up .29 and .31, respectively) (NICE, 2013). However, this finding needs to be considered with caution because only five studies were available for this IPD-MA. In addition, the sample included in this IPD-MA was heterogeneous, with studies including participants recruited from different settings (i.e., general population, occupational setting, and



specialized mental health care) different depression severity (i.e., major depression diagnosis and subclinical depression), and comorbid conditions (i.e., diabetes).

The total of six eligible studies (we could not access the data from one study) identified in this IPD-MA is only a subset of the published RCTs that examined the clinical effectiveness of guided Internet-based interventions. For instance, a recent meta-analysis identified 18 eligible studies that compared depression outcomes between guided Internet-based interventions for depression in and control conditions (Ebert et al., 2016). Thus, only a minority of studies evaluating guided Internet-based interventions for depression included a cost-effectiveness analysis alongside the effectiveness analysis. It appears that this is a selected subset of studies, since previous meta-analyses examining the effectiveness of guided Internet-based interventions compared to control groups have found moderate to large effect sizes (Cohen's *d* between .61 and 1.00) (Andersson & Cuijpers, 2009; Spek et al., 2007). In contrast, the differences in effectiveness between the intervention and control in the studies included in our IPD-MA is smaller.

Our results are not in line with findings from earlier systematic reviews evaluating the cost effectiveness of Internet-based treatments (Donker et al., 2015; Hedman et al., 2012). The authors of previous studies concluded that Internet-based treatments have high probability of being cost effective compared to controls for various mental health problems including depression (Donker et al., 2015; Hedman et al., 2012). Other authors, however, stated that the information they had was not sufficient to draw firm conclusions (Andersson, Wagner, & Cuijpers, 2016; Arnberg, Linton, Hultcrantz, Heintz, & Jonsson, 2014). A possible explanation for the difference between our conclusions and those of some of the previous studies is the absence of guidelines on what probability of cost effectiveness is perceived acceptable by healthcare decision makers. As a result, sometimes relatively low probabilities (e.g., anything higher than .50) have led to the conclusion that the intervention under study was cost effective in comparison with its comparator (Hedman et al., 2012). In addition, systematic reviews typically do not pool data from cost-effectiveness studies and researchers are limited to descriptive analysis of point estimates of the ICERs without assessing statistical uncertainty.

It has been hypothesized that Internet-based interventions are cost effective compared to face-to-face treatments due to the reduced time the therapists spend with patients. However, this hypothesis was not confirmed in the present study where the intervention was more expensive than control, even though not statistically significantly so. A potential explanation for this is that patients in the intervention group received the guided Internet-based intervention on top of the face-to-face sessions with healthcare professionals, resulting in increased costs. Another explanation may be that patients in the control groups did not always receive a structured, manualized treatment, and, thus, had fewer face-to-face sessions with healthcare professionals, resulting in lower costs. This, for example, was the case in the study by Warmerdam et al. (2010), in which the comparator was a waiting list group. The importance of including the appropriate control group in comparative research has been discussed before and it has been highlighted that a thorough description of the treatment the patients in the

control group receive is necessary (Kolovos et al., 2017; Watts, Turnell, Kladnitski, Newby, & Andrews, 2015).

A recent meta-analysis has shown that psychological treatments are more effective than control conditions to improve QoL of patients with depression (Kolovos, Kleiboer, & Cuijpers, 2016). However, the present study did not find any clinically relevant differences in QALYs between the two groups. A possible explanation is that the EQ-5D-3L, which was used to calculate QALYs in this study, is not sensitive enough to detect small changes in QoL as a result of depression treatment (Brazier et al., 2014).

We conducted eight sensitivity analyses to examine the robustness of the findings of the main analyses. The results from most sensitivity analyses were in line with our main findings. The results from the sensitivity analysis including participants recruited from the general population (SA7), demonstrated a high probability in relatively low ceiling ratios for cost effectiveness of the intervention compared to control for depressive symptoms but only moderate probability for cost effectiveness when QALYs were considered. Because only two studies ( $n = 663$ ) were included in these analyses, we cannot draw firm conclusions.

The current study has several strengths. To our knowledge, this is the first study that uses IPD-MA to evaluate the cost effectiveness of guided Internet-based depression interventions compared to control. This novel approach allowed us to overcome obstacles that were encountered by previous studies, such as different follow-up measurements and differences in statistical approaches. Furthermore, even though only a few studies that investigated the cost effectiveness of guided Internet-based interventions for depression were available, pooling the IPD resulted in a much larger sample size than commonly used in economic evaluations conducted alongside clinical trials. This is particularly important considering the skewed distribution of costs, which implies that much larger samples are needed (Briggs, 2000).

There are also some study limitations that we need to consider. First, we collected data from five of the six eligible studies. The findings from the study that we were not able to include were more favorable for the guided Internet-based intervention as compared to the findings of the current IPD-MA (Titov et al., 2015). The ICER for QALYs was \$4,392, with a > 95% probability of the intervention being cost effective at a WTP threshold of \$50,000/QALY, which is a commonly used threshold in Australia, where the study was conducted. However, it should be considered that the aforementioned study included a waiting list group as a comparator, which is not considered an optimal comparator in an economic evaluation. Moreover, we used the Dutch tariffs to calculate the utility weights for EQ-5D-3L. The utility weights could be slightly different if we had used tariffs generated from a different population (Badia, Roset, Herdman, & Kind, 2001). However, a recent study examining the mean utility scores for different health states related to depression found that the tariffs used (i.e., United Kingdom and Dutch) did not considerably influence the results (Kolovos et al., 2017).

Overall, to the best of our knowledge, this is the first study that pooled cost and effect estimates from different clinical trials in an IPD-MA evaluating the cost effectiveness of Internet-based interventions. Based on our results, the Internet-based intervention cannot be considered cost effective compared to control for depression severity,

treatment response, and QALYs. However, these results should be interpreted with caution since only a minority of RCTs investigating the clinical effectiveness of guided Internet-based interventions also examined cost effectiveness and were included in this IPD-MA. We strongly recommend adding a cost-effectiveness analysis to future RCTs.

## ACKNOWLEDGMENTS

The current study has been conducted within the E-COMPARED framework (603098). The E-COMPARED project is funded under the Seventh Framework Program. The content of this article reflects only the authors' views and the European Community is not liable for any use that may be made of the information contained therein.

## CONFLICT OF INTEREST

The authors report no conflict of interest.

## ORCID

Spyros Kolovos MSc  <http://orcid.org/0000-0003-3201-1743>

Pim Cuijpers PhD  <http://orcid.org/0000-0001-5497-2743>

## REFERENCES

- Andersson, G., & Cuijpers, P. (2009). Internet-based and other computerized psychological treatments for adult depression: A meta-analysis. *Cognitive Behaviour Therapy*, 38(4), 196–205. <https://doi.org/10.1080/16506070903318960>
- 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(3), 288–295. <https://doi.org/10.1002/wps.20151>
- Andersson, G., Topooco, N., Havik, O., & Nordgreen, T. (2016). Internet-supported versus face-to-face cognitive behavior therapy for depression. *Expert Review of Neurotherapeutics*, 16(1), 55–60. <https://doi.org/10.1586/14737175.2015.1125783>
- Andersson, G., Wagner, B., & Cuijpers, P. (2016). ICBT for depression. In N. Lindefors & G. Andersson (Eds.), *Guided Internet-based treatments in psychiatry* (pp. 17–32). Cham: Springer International Publishing.
- Arnberg, F. K., Linton, S. J., Hultcrantz, M., Heintz, E., & Jonsson, U. (2014). Internet-delivered psychological treatments for mood and anxiety disorders: A systematic review of their efficacy, safety, and cost-effectiveness. *PLoS One*, 9(5), 1–13. <https://doi.org/10.1371/journal.pone.0098118>
- Badia, X., Roset, M., Herdman, M., & Kind, P. (2001). A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Medical Decision Making*, 21(1), 7–16.
- Bouwman, C., De Jong, K., Timman, R., Zijlstra-Vlasveld, M., Van der Feltz-Cornelis, C., Tan, S. S., & Hakkaart-van Roijen, L. (2013). Feasibility, reliability and validity of a questionnaire on health-care consumption and productivity loss in patients with a psychiatric disorder (TIC-P). *BMC Health Services Research*, 13, 1–9. <https://doi.org/10.1186/1472-6963-13-217>
- Brazier, J., Connell, J., Papaioannou, D., Mukuria, C., Mulhern, B., Peasgood, T., ... Knapp, M. (2014). A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. Health technology assessment (Winchester, England), 18(34), vii.
- Briggs, A. (2000). Economic evaluation and clinical trials: Size matters. *BMJ*, 321(7273), 1362–1363.
- Buntrock, C., Berking, M., Smit, F., Lehr, D., Nobis, S., Riper, H., ... Ebert, D. (2017). Preventing depression in adults with subthreshold depression: Health-economic evaluation alongside a pragmatic randomized controlled trial of a web-based intervention. *Journal of Medical Internet Research*, 19(1), 1–16. <https://doi.org/10.2196/jmir.6587>
- Donker, T., Blankers, M., Hedman, E., Ljotsson, B., Petrie, K., & Christensen, H. (2015). Economic evaluations of Internet interventions for mental health: A systematic review. *Psychological Medicine*, 45(16), 3357–3376. <https://doi.org/10.1017/s0033291715001427>
- Ebert, D. D., Donkin, L., Andersson, G., Andrews, G., Berger, T., Carlbring, P., ... Cuijpers, P. (2016). 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. *Psychological Medicine*, 46(13), 2679–2693. <https://doi.org/10.1017/s0033291716001562>
- Efron, B. (1994). Missing data, imputation, and the bootstrap. *Journal of the American Statistical Association*, 89(426), 463–475.
- EuroQol-Group. (1990). EuroQol—A new facility for the measurement of health-related quality of life. *Health Policy*, 16(3), 199–208.
- Evers, S. M., Hilgsmann, M., & Adarkwah, C. C. (2015). Risk of bias in trial-based economic evaluations: Identification of sources and bias-reducing strategies. *Psychology & Health*, 30(1), 52–71. <https://doi.org/10.1080/08870446.2014.953532>
- Faria, R., Gomes, M., Epstein, D., & White, I. R. (2014). A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. *Pharmacoeconomics*, 32(12), 1157–1170.
- Fenwick, E., O'Brien, B. J., & Briggs, A. (2004). Cost-effectiveness acceptability curves—Facts, fallacies and frequently asked questions. *Health Economics*, 13(5), 405–415.
- Geraedts, A. S., van Dongen, J. M., Kleiboer, A. M., Wiezer, N. M., van Mechelen, W., Cuijpers, P., & Bosmans, J. E. (2015). Economic evaluation of a web-based guided self-help intervention for employees with depressive symptoms: Results of a randomized controlled trial. *Journal of Occupational and Environmental Medicine*, 57(6), 666–675. <https://doi.org/10.1097/jom.0000000000000423>
- Hedman, E., Ljotsson, B., & Lindefors, N. (2012). Cognitive behavior therapy via the Internet: A systematic review of applications, clinical efficacy and cost-effectiveness. *Expert Review of Pharmacoeconomics & Outcomes Research*, 12(6), 745–764. <https://doi.org/10.1586/erp.12.67>
- Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., ... Sterne, J. A. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 343, 1–9.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12–19.
- Karyotaki, E., Kleiboer, A., Smit, F., Turner, D. T., Pastor, A. M., Andersson, G., ... Cuijpers, P. (2015). Predictors of treatment dropout in self-guided web-based interventions for depression: An “individual patient data” meta-analysis. *Psychological Medicine*, 45(13) 2717–2726. <https://doi.org/10.1017/s0033291715000665>
- Kessler, R. C., & Bromet, E. J. (2013). The epidemiology of depression across cultures. *Annual Review of Public Health*, 34, 119–138. <https://doi.org/10.1146/annurev-publhealth-031912-114409>
- Kessler, R. C., Sampson, N. A., Berglund, P., Gruber, M., Al-Hamzawi, A., Andrade, L., ... De Girolamo, G. (2015). Anxious and non-anxious major depressive disorder in the World Health Organization World Mental Health Surveys. *Epidemiology and Psychiatric Sciences*, 24(03), 210–226.

- Kolovos, S., Bosmans, J. E., van Dongen, J. M., van Esveld, B., Magai, D., van Straten, A., ... van Tulder, M. W. (2017). Utility scores for different health states related to depression: Individual participant data analysis. *Quality of Life Research*, 26, 1649–1658. <https://doi.org/10.1007/s11136-017-1536-2>
- Kolovos, S., Kenter, R. M., Bosmans, J. E., Beekman, A. T., Cuijpers, P., Kok, R. N., & van Straten, A. (2016). Economic evaluation of Internet-based problem-solving guided self-help treatment in comparison with enhanced usual care for depressed outpatients waiting for face-to-face treatment: A randomized controlled trial. *Journal of Affective Disorders*, 200, 284–292. <https://doi.org/10.1016/j.jad.2016.04.025>
- Kolovos, S., Kleiboer, A., & Cuijpers, P. (2016). The effect of psychotherapy for depression on quality of life: A meta-analysis. *British Journal of Psychiatry*, 209, 1–9.
- Kolovos, S., van Tulder, M. W., Cuijpers, P., Prigent, A., Chevreur, K., Riper, H., & Bosmans, J. E. (2017). The effect of treatment as usual on major depressive disorder: A meta-analysis. *Journal of Affective Disorders*, 210, 72–81. <https://doi.org/10.1016/j.jad.2016.12.013>
- Lamers, L. M., Stalmeier, P. F., McDonnell, J., Krabbe, P. F., & van Busschbach, J. J. (2005). Measuring the quality of life in economic evaluations: The Dutch EQ-5D tariff. *Nederlands Tijdschrift voor Geneeskunde*, 149(28), 1574–1578.
- Lim, D., Sanderson, K., & Andrews, G. (2000). Lost productivity among full-time workers with mental disorders. *Journal of Mental Health Policy and Economics*, 3(3), 139–146.
- Maldonado, G., & Greenland, S. (1993). Simulation study of confounder-selection strategies. *American Journal of Epidemiology*, 138(11), 923–936.
- NICE. (2013). *Guide to the methods of technology appraisal*. Retrieved from <https://nice.org.uk/process/pmg9>
- Nobis, S., Lehr, D., Ebert, D. D., Berking, M., Heber, E., Baumeister, H., ... Riper, H. (2013). Efficacy and cost-effectiveness of a web-based intervention with mobile phone support to treat depressive symptoms in adults with diabetes mellitus type 1 and type 2: Design of a randomised controlled trial. *BMC Psychiatry*, 13, 776–783.
- Ramsey, S., Willke, R., Briggs, A., Brown, R., Buxton, M., Chawla, A., ... Reed, S. (2005). Good research practices for cost-effectiveness analysis alongside clinical trials: The ISPOR RCT-CEA Task Force report. *Value Health*, 8(5), 521–533. <https://doi.org/10.1111/j.1524-4733.2005.00045.x>
- Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: Rationale, conduct, and reporting. *BMJ*, 340, 1–7. <https://doi.org/10.1136/bmj.c221>
- Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys* (Vol. 81). Hoboken, New Jersey: John Wiley & Sons.
- Shim, R. S., Baltrus, P., Ye, J., & Rust, G. (2011). Prevalence, treatment, and control of depressive symptoms in the United States: Results from the National Health and Nutrition Examination Survey (NHANES), 2005–2008. *Journal of the American Board of Family Medicine*, 24(1), 33–38. <https://doi.org/10.3122/jabfm.2011.01.100121>
- Simons, C. L., Rivero-Arias, O., Yu, L. M., & Simon, J. (2015). Multiple imputation to deal with missing EQ-5D-3L data: Should we impute individual domains or the actual index? *Quality of Life Research*, 24(4), 805–815.
- Spek, V., Cuijpers, P., Nyklicek, I., Riper, H., Keyzer, J., & Pop, V. (2007). Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: A meta-analysis. *Psychological Medicine*, 37(3), 319–328.
- Statistics-Nederlands. (2016). Voorburg/Heerlen. From Centraal Bureau voor de Statistiek (CBS).
- Stewart, L. A., & Parmar, M. K. (1993). Meta-analysis of the literature or of individual patient data: Is there a difference? *Lancet*, 341(8842), 418–422.
- Tate, D. F., Finkelstein, E. A., Khavjou, O., & Gustafson, A. (2009). Cost effectiveness of internet interventions: Review and recommendations. *Annals of Behavioral Medicine*, 38(1), 40–45. <https://doi.org/10.1007/s12160-009-9131-6>
- Titov, N., Dear, B. F., Ali, S., Zou, J. B., Lorian, C. N., Johnston, L., ... Fogliati, V. J. (2015). Clinical and cost-effectiveness of therapist-guided internet-delivered cognitive behavior therapy for older adults with symptoms of depression: A randomized controlled trial. *Behavior Therapy*, 46(2), 193–205.
- Van Buuren, S., Boshuizen, H. C., & Knook, D. L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine*, 18(6), 681–694.
- Vos, T., Flaxman, A. D., Naghavi, M., Lozano, R., Michaud, C., Ezzati, M., ... Memish, Z. A. (2012). Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, 380(9859), 2163–2196. [https://doi.org/10.1016/s0140-6736\(12\)61729-2](https://doi.org/10.1016/s0140-6736(12)61729-2)
- Vroomen, J. M., Eekhout, I., Dijkgraaf, M. G., van Hout, H., de Rooij, S. E., Heymans, M. W., & Bosmans, J. E. (2015). Multiple imputation strategies for zero-inflated cost data in economic evaluations: Which method works best? *The European Journal of Health Economics*, 17(8), 939–950.
- Warmerdam, L., Smit, F., van Straten, A., Riper, H., & Cuijpers, P. (2010). Cost-utility and cost-effectiveness of Internet-based treatment for adults with depressive symptoms: Randomized trial. *Journal of Medical Internet Research*, 12(5), 40–50.
- Watts, S. E., Turnell, A., Kladnitski, N., Newby, J. M., & Andrews, G. (2015). Treatment-as-usual (TAU) is anything but usual: A meta-analysis of CBT versus TAU for anxiety and depression. *Journal of Affective Disorders*, 175, 152–167. <https://doi.org/10.1016/j.jad.2014.12.025>
- White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4), 377–399. <https://doi.org/10.1002/sim.4067>
- Zwaap, J., Knies, S., van der Meijden, C., Staal, P., & van der Heiden, L. (2015). Kosteneffectiviteit in de praktijk. In Z. Nelderland (Ed.), 1–78. Diemen: Ministerie van VWS.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Kolovos S, van Dongen JM, Riper H, et al. Cost effectiveness of guided Internet-based interventions for depression in comparison with control conditions: An individual-participant data meta-analysis. *Depress Anxiety*. 2018;35:209–219. <https://doi.org/10.1002/da.22714>