Internet-based treatment for eating disorders: bridging the treatment gap

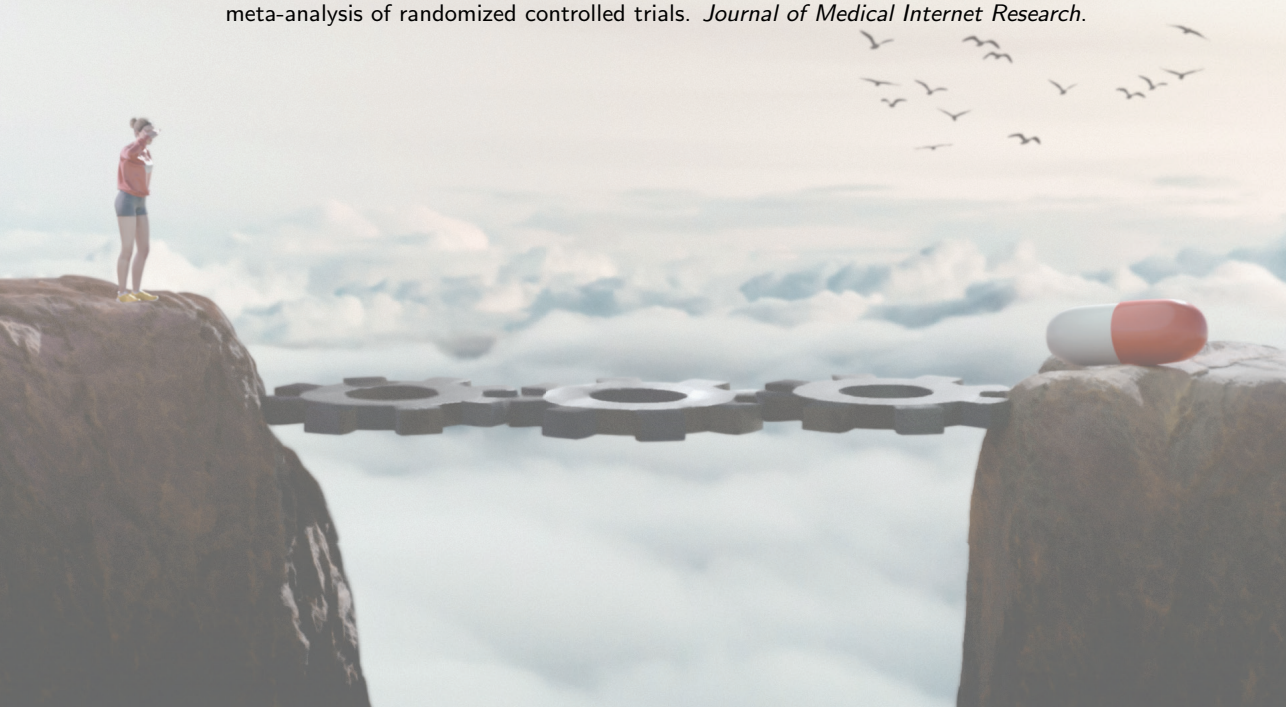
Rohrbach, P.J.

Chapter 4

Cost-effectiveness of internet interventions compared to treatment as usual for people with mental disorders: a systematic review and meta-analysis of randomized controlled trials

Trial Registration: Prospero registration https://www.crd.york.ac.uk/PROSPERO/; registration number: CRD42019141659.

Accepted for publication: Rohrbach, P. J., Dingemans, A. E., Evers, C. A. J. M., Aardoom, J. J., Van Furth, E. F., Spinhoven, P., Lähde, I., Clemens, F., & Van den Akker-Van Marle, M. E. (in press). Cost-effectiveness of eHealth interventions compared to treatment as usual for people with mental disorders: a systematic review and meta-analysis of randomized controlled trials. *Journal of Medical Internet Research*.



Abstract

Background: Economic costs of mental disorders for society are huge. Internet-based interventions are often coined as cost-effective alternative to usual care, but evidence is mixed.
Objective: The aim was to review the literature on cost-effectiveness of internet interventions for mental disorders compared to usual care and to provide an estimate of the monetary benefits of such interventions compared to usual care. **Methods:** A systematic review and meta-analysis of (1) randomized controlled trials that (2) included participants with symptoms of mental disorders, (3) investigated a telephone or internet-based intervention, (4) included a control condition in the form of treatment as usual, psychological placebo, waiting list control or bibliotherapy, (5) reported outcomes on both quality of life and costs and (6) were published in English, was conducted. Electronic databases PubMed (including Medline), Embase, Emcare, PsycINFO, Web of Science and The Cochrane Library were used. Data on risk of bias, quality of the economic evaluation, quality adjusted life years (QALYs) and costs were extracted from included studies and the incremental net benefit (INB) was calculated and pooled.
Results: The search yielded 6226 abstracts and 37 studies with 14,946 participants were included. Quality of economic evaluations of included studies was rated to be moderate and risk of bias was high. A random-effects approach was maintained. Analyses suggested internet interventions to be slightly more effective than usual care in terms of QALY gain, Hedges' $g = 0.05$ (95% CI 0.01; 0.10, $P = .016$), and equally expensive, Hedges' $g < 0.01$ (95% CI -0.08; 0.84, $P = .96$). The pooled INB was \$255, (95% CI \$91; \$419, $P = .002$), favoring eHealth interventions over usual care. Perspective of the economic evaluation and targeted mental disorder moderated results.
Conclusions: Findings indicate that cost-effectiveness of internet interventions for mental disorders compared to a care-as-usual approach is likely, but generalizability to new studies is poor given the substantial heterogeneity. This is the first study in the area of mental health to pool cost-effectiveness outcomes in an aggregate-data meta-analysis.

Introduction

Mental disorders have a huge impact on sufferers as well as on society. It is appraised to cause almost one-third of global years lived with disability and account for roughly 10% of all disability-adjusted life years, placing it in the top 3 causes of global burden worldwide (Vigo et al., 2016). When mental, neurological and substance use disorders are taken together, the global economic costs in 2010 were estimated to be 2.5 trillion US dollars, with projections for 2030 being around 6 trillion dollars (Marquez & Saxena, 2016). While these numbers are serious and account for around 2.3-4.4% of the gross domestic product in high-income countries, most countries spend a disproportionally small amount of their health budget on mental health (Organisation for Economic Co-operation and Development, 2014). This warrants changes in policy, but also stresses the need for effective and inexpensive interventions so that individuals can recover more swiftly from a mental disorder and the global burden is reduced.

Effectiveness of internet interventions

Swift technological advancement brings the promise of new and effective interventions. Indeed, internet interventions have become a popular niche of research and treatment. A reason for attempting to create effective internet interventions is to reduce the treatment gap, which specifies the discrepancy between the proportion of people who need help for a particular disorder and the proportion of those individuals who actually receive care (Bennett & Glasgow, 2009; Kazdin et al., 2017). In other words, internet interventions can be used to reach an underserved population of people with a (risk for developing) mental illness (Aardoom, Dingemans, & Van Furth, 2016). Internet interventions for people with (symptoms of) a mental disorder are increasingly confirmed in their effectiveness. There have been several meta-analyses on this topic covering various mental disorders, such as depression, (social) anxiety, post-traumatic stress and eating disorders (Andersson et al., 2014; Carlbring et al., 2017; Melioli et al., 2016; Sander et al., 2016; Välimäki et al., 2017). A recent umbrella review of meta-analyses shows that there is sufficient information now to assume the effectiveness of internet interventions (Andersson et al., 2019). In general, guided internet interventions (most of which have a cognitive-behavioral underpinning) seem to be as effective in reducing symptomatology as face-to-face treatment and outperform waiting list control conditions. Unguided internet interventions seem to be more effective than waiting list control conditions, but less effective than guided internet or face-to-face interventions.

Cost-effectiveness of internet interventions

Internet interventions for mental disorders are often coined as a cost-effective alternative to established treatments (F. Griffiths et al., 2006), but results on cost-effectiveness are tentative at best (Hedman et al., 2012; Tate et al., 2009). Individual studies show mixed results and their heterogeneity in methods, outcomes and comparators makes it difficult to draw conclusions (Donker et al., 2015; Naslund et al., 2022), so no definitive assumptions can yet be established. Additionally, cost-effectiveness studies that are conducted alongside randomized controlled trials (RCTs) are often not powered for economic evaluations, limiting

their predictive ability (Hollingworth et al., 2013). Nevertheless, the separate studies make an important contribution to the rapidly growing body of evidence, so that it becomes more feasible to make meaningful overviews in the form of systematic reviews and meta-analyses. Indeed, Naslund et al. (2022) performed a broad systematic review on cost outcomes for telemedicine interventions for mental disorders and Donker et al. (2015) systematically reviewed RCTs on cost-effectiveness of internet-based interventions for mental disorders. Both reviews concluded that internet interventions for mental disorders have the potential to be cost-effective compared to alternatives, but evidence is still circumstantial. Kolovos et al. (2018) performed an individual participant meta-analysis to investigate the cost-effectiveness of internet interventions compared to a control condition (e.g., waiting list or care as usual) for depression. The authors cautiously conclude that results showed no indication of cost-effectiveness and remark that adding economic evaluations to trials more frequently would help to reach well-founded deductions.

Pooling cost-effectiveness data

To get a precise estimate of the cost-effectiveness of internet interventions for mental disorders compared to alternatives it is desirable to pool outcomes of individual studies in an aggregate data meta-analysis. As cost-effectiveness is expressed as a combination of two variables, the difference in costs and effects between an intervention and control condition, this is statistically complex. Fortunately, Crespo et al. (2014) developed a theoretical framework to do these kinds of analyses. This method has been successfully applied by another research team in several studies in different areas of medicine (Bagepally et al., 2020; Bagepally et al., 2019; Chaiyakittisopon et al., 2021; Haider et al., 2019; Noparatayaporn et al., 2021). Currently, no meta-analysis on the cost-effectiveness of internet interventions for mental disorders has been conducted. The individual participant data (IPD) meta-analysis by Kolovos et al. (2018) looked at depression only. The IPD approach is more reliable (Riley et al., 2010), but decreasingly feasible when more studies are included. Furthermore, the last overview of RCTs on cost-effectiveness of internet-based interventions for mental disorders was in 2015 (Donker et al., 2015), but results were not pooled in a meta-analysis.

Concordantly, given the mixed evidence and limitations of individual studies and the rapid increase of novel research, a thorough and recent overview of the literature on the cost-effectiveness of internet interventions for individuals with (symptoms of) a mental disorder compared to alternatives is warranted. A first step would be to establish whether internet interventions are cost-effective compared to control conditions such as a waiting list and care as usual. The current article, then, aims to investigate the cost-effectiveness of internet interventions for mental disorders compared to control conditions by conducting a systematic literature search and aggregate data meta-analysis of RCTs.

Methods

The literature review was conducted according to the Preferred Reporting Items for Systematic Review (PRISMA) guidelines (Moher et al., 2015), see Appendix C.1. This review was registered in Prospero (registration number CRD42019141659).

Eligibility criteria

Studies were included that (1) were RCTs; (2) included participants (of all ages) who have symptoms of or a diagnosed mental disorder; (3) investigated a telephone- or an internet-based (that work via computer, tablet and/or smartphone) intervention. All forms of internet-based interventions were considered, including fully automated (i.e., unguided) interventions, guided interventions and teleconferencing interventions. Guided interventions could contain asynchronous support (i.e., a delay in the support such as with mail or forum services) or synchronous support (i.e., no delay in the support such as with videoconferencing and chat). Additionally, smartphone apps with the purpose of elevating symptoms of a mental disorder were included; (4) included a control condition in the form of (enhanced) treatment as usual, psychological placebo, waiting list control or bibliotherapy; (5) reported outcomes on both quality of life (quality adjusted life years; QALYs) and costs; (6) were published in English. It is worth noting that studies were included if participants had somatic conditions (e.g., cancer or diabetes) as long as the participants had comorbid symptoms of a mental disorder and the investigated intervention had as primary aim to alleviate these symptoms. Studies were excluded if the main intervention exclusively relied on wearable devices or virtual reality, had only a face-to-face intervention as control condition, or did not provide sufficient information on costs or QALYs.

Search strategy and selection criteria

Electronic databases PubMed (including Medline), Embase, Emcare, PsycINFO, Web of Science and The Cochrane Library were searched up until 1 March 2021. A search string was made for PubMed and translated for the other databases. The PubMed search string contained MeSH terms for the concepts of 'mental disorders', 'cost-benefit analysis' and 'telemedicine'/'internet', with all expressions under these headings included as free terms as well. Furthermore, other terms related to the three MeSH terms were added to maximize the sensitivity of the search. The full PubMed search strategy can be found in the Appendix C.2. By checking cross-references in the included studies we minimized the chance of missing relevant data. During the screening phase, all relevant study protocols, conference abstracts and trial registrations were identified and authors of unpublished studies with potentially relevant data were contacted.

The identified articles were all screened in three steps and by two researchers (PR, AD, CE, EA, IL or FC). First, the title and abstract of the eligible articles were screened. Subsequently, the full texts of all the included abstracts were screened for eligibility. Lastly, the relevant data were extracted. If there was any disagreement between the two researchers in any of the three steps a third researcher of the team made the final decision.

Quality assessment

Risk of bias of the included RCTs was assessed by two researchers (PR, AD, CE, IL or FC) using the Cochrane Risk of Bias assessment tool (Higgins et al., 2011). Specifically, data were gathered on the topics of (1) random sequence generation (selection bias), (2) allocation concealment (selection bias), (3) blinding of participants and personnel (performance bias),

(4) blinding of outcome assessment (detection bias), (5) incomplete outcome data (attrition bias), (6) selective reporting (reporting bias), (7) other sources of bias. Finally, for all these topics the risk of bias was assessed (i.e., low, unclear or high risk of bias). Any disagreement between the researchers on (each of the seven areas of) the risk of bias was resolved by means of a discussion between the two raters or by a third rater. A final rating of high, medium or low bias was assigned to each study based on the revised tool for assessing risk of bias (Sterne et al., 2019). Specifically, high risk of bias was assigned to studies when a high risk of bias in any domain was present or unclear risk of bias was present in two or more domains; medium risk of bias was assigned in the case of one unclear rating across all domains; low risk of bias was assigned when all domains were rated as low risk of bias. Exceptions were that blinding of participants and personnel was not considered in the final risk of bias rating as blinding was unfeasible in most studies, and unclear risk of bias on the selective reporting domain was considered as low risk of bias for the final risk of bias rating as the absence of a published protocol or trial preregistration is still common.

Furthermore, the 19-item CHEC list (Evers et al., 2005) was used to assess the quality of the economic evaluation for all included studies. The expert on economic evaluations (EA) scored all articles on the CHEC list, while other authors (PR, AD, CE, IL and FC) divided included articles for the CHEC list as well, so articles were rated twice. All discrepancies between raters on the CHEC list were resolved by means of a discussion. Assigning an overall quality score for the CHEC list is not advocated, as cutoff scores are highly heterogeneous (Watts & Li, 2019) and difficult to substantiate. Since we wanted to group studies by quality of the economic evaluation in analyses, we adopted an approach using a selection of items of the CHEC list. A study had to fulfill at least 8 specific items (i.e., items 7-14, see Table 4 for a list of all items) to be deemed of high quality. These items were chosen, because they contribute to the assessment of cost and effects used in the current meta-analysis. Other items (i.e., 1-4 and 15-19) were deemed less important for this study, as they aim at clarifying the text and the appropriateness of the used methods and analyses, which is also captured partly in the risk of bias assessment. Finally, items 5 and 6 regarding time horizon and perspective were not considered for the final quality rating, because these were already explored in moderator analyses.

Outcome measures

Quality of life

The difference in the average number of QALYs gained per participant in the intervention group and its comparator (delta QALY) was the target health outcome measure. A QALY indicates one extra life year in perfect health. QALYs are derived from generic health-related quality-of-life measures (e.g. EQ-5D or 36-item Short Form Health Survey; SF-36) that are transformed into utility scores, multiplied by the time (in years) a participant spent with that utility (Singh et al., 2001).

Costs

Delta societal costs or delta healthcare costs (from now on referred to as delta costs), which indicate the difference in costs between the intervention and control condition, was the

primary outcome measure. Studies with a healthcare perspective estimated costs per study group by measuring healthcare use of participants during the intended follow-up period and multiplying that with a reference price for the used services. Studies with a societal perspective also included costs based on, for example, absence from paid and unpaid work, reduced productivity while at work (presenteeism) and/or domestic care for participants.

Two steps were taken to account for cost differences. First, inflation was controlled for by recalculating all costs to the level of 2021 using consumer price indexes as reported for the countries in which each study was conducted (OECD, 2021b). Second, similar articles or services have different prices between countries, indicating differences in purchasing power. Purchasing Power Parities (PPP) were used to transform costs according to the most recent rates from 2017 (OECD, 2021a), effectively converting all costs to US dollars. A general indexing factor was used for PPP rather than a healthcare specific one, since other costs were also involved when studies employed a societal perspective.

Data preparation

All data extraction items can be found in the Appendix C.3. Apart from the outcome measures needed for the meta-analysis, other data were extracted either for sensitivity analyses or for further exploration. Data needed for the meta-analysis were (1) delta QALY, (2) variance of delta QALY, (3) delta costs, (4) variance of delta costs, and (5) covariance of delta QALY and delta costs. With these variables available for each study it was possible to calculate the incremental net benefit (INB) and pool all INBs using the method as described by Crespo et al. (2014). The INB indicates gains of an intervention compared to another expressed in monetary terms. Delta QALY and delta costs were either retrieved directly from the articles or calculated by subtracting the average QALYs or costs per patient in the intervention condition from those in the control condition. The method to retrieve the variances and covariance was based on that of Bagepally et al. (2019) and described here in order of the most ideal (i.e., reliable) to least ideal scenario. Appendix C.4 lists all used formulas for this meta-analysis including the INB.

Calculating the variance of delta QALY

Scenario 1. Studies report the standard deviation (SD), standard error (SE) or 95% confidence interval (CI) of delta QALY. The variance can directly be calculated by formula 1 or 2. When the 95% CI is reported an additional step is required, for which we used formula 3.

Scenario 2. Studies report the SD, SE or 95% CI of QALYs for the separate conditions (but not for the difference). If the SDs of separate conditions are reported the variance of the difference between two conditions can be calculated using formula 4. If the SEs of separate conditions are reported the variance can be calculated using formula 5. If the 95% CIs of separate conditions are reported the SE is first calculated using formula 3.

Scenario 3. Studies report only the average of separate conditions, but no measure of spread. In this case, first the corresponding author of the article was contacted to inquire about the possibility of receiving an indication of spread (SD, SE or 95% CI) of delta QALY or of the spread of the average QALY gain per condition. If this was not possible or the authors did not respond, two further options remained to estimate the measure of spread.

First, if a cost-utility plane with bootstrapped delta QALY and delta healthcare or societal costs was reported, the free available software Webplot Digitizer (Rohatgi, 2020) was used to reverse engineer the individual data points. The points were exported to an Excel file, from which the SD of delta QALY could be calculated. An example of how the software was used can be found in the Appendix C.5. Webplot Digitizer has been used in various studies and is found to be a reliable tool for extracting data with high intercoder reliability and validity (Burda et al., 2017; Drevon et al., 2017). However, the precision of the software seems to be dependent on the visual presentation of individual graphs (Moeyaert et al., 2016). Second, if no cost-effectiveness plane was reported the measure of spread of delta QALY was estimated by taking the mean of the two most similar studies or comparisons in terms of delta QALY, number of participants and investigated intervention.

Calculating the variance of delta costs

For these calculations identical steps were followed as for the variance of delta QALY, but using costs instead of QALYs.

Calculating the covariance between delta QALY and delta costs

To estimate the covariance between delta QALY and delta costs a bootstrap procedure is necessary to be able to have multiple estimates of delta QALY and delta costs. The covariance was never reported in included articles. As shown by (Bagepally et al., 2019), an approximation without the original data is possible, albeit less reliable. Hence, all corresponding authors of included studies were contacted and asked to provide the covariance, to rely on author data for estimates of this parameter as much as possible.

Scenario 1. Authors were able to provide the covariance or information needed to calculate it (e.g., the data set including all bootstrapped delta QALYs and delta costs or individual participant data needed to perform a bootstrap procedure including (1) the allocated condition of participants, (2) the QALYs over the entire follow-up period and (3) the costs over the entire follow-up period).

Scenario 2. If authors did not respond or could not provide information on the covariance, but a cost-utility plane was presented in the article Webplot Digitizer was used. The covariance could be calculated after using the software to estimate the data points (delta QALYs and delta costs).

Scenario 3. If authors did not respond or could not provide information on the covariance and no cost-utility plane was presented in the article, a different approach was used. First, the mean correlation between delta QALY and delta costs was calculated for all studies where the covariance was obtained with data received directly from authors (i.e., covariances obtained from scenario 1) using formula 6. Second, the mean correlation was used to calculate the covariance between delta QALY and delta costs for studies falling under scenario 3 using formula 6 (mean imputation).

Statistical analyses

Data preparation was done in Excel and analyses were conducted in Comprehensive Meta-Analysis software, version 3 (Borenstein et al., 2013). First, we calculated the INB for each study. The first step is to multiply society's willingness to pay (WTP) for one extra year lived in perfect health (i.e., 1 QALY) with the difference in effectiveness (delta QALYs) between two interventions (internet intervention versus control). This expresses the difference in effects in monetary terms. Subtracting the difference in costs between the two conditions (delta costs) results in the INB (see formula 7). WTP for one QALY was set at $40,000 dollars. This value was based on WTP values in high-income countries, which are typically around $40,000 per QALY (Schwarzer et al., 2015).

Pooling incremental net benefits

As studies were expected to be heterogeneous concerning follow-up periods, included costs and sampled population, a random effects approach was maintained throughout the analyses, regardless of heterogeneity scores like the Cochran Q and I-squared. In accordance with this approach, study weights were corrected based on the DerSimonian and Laird method (DerSimonian & Laird, 1986).

Moderators

Several moderators were incorporated in the analyses to explore their influence on the overall cost-effectiveness. Specifically, subgroups were based on (1) the perspective of the economic evaluation (health care versus societal), (2) length of follow-up (12 months or longer versus shorter than 12 months) and (3) targeted mental disorder. Other considered subgroups were based on (4) presence of guidance (yes/no), (5) intensity of the guidance (self-guided, less than weekly, weekly or more than weekly), (6) type of guidance (asynchronous/delayed such as e-mail, synchronous/immediate such as chat or telephone, or a combination), (7) method of recruitment (open/mass media or clinical referral), (8) method of diagnosis for inclusion (formal diagnosis or self-reported symptoms), (9) duration of the intervention (4-8 weeks, 9-12 weeks, longer than 12 weeks or undefined/unlimited access), and (10) type of control condition (care as usual or attention control).

Heterogeneity and publication bias

To get an impression of the heterogeneity between studies I^2 and Cochran Q (formulas 12 and 13) were calculated and reported for all analyses. However, visual inspection of the forest plot and consideration of the study characteristics were leading in the identification of between-study heterogeneity. Indications of publication bias were explored with both a visual inspection of the funnel plot and the Eggers test.

Sensitivity analyses

Robustness of the results were inspected in five sensitivity analyses. Specifically, the pooled INB was calculated separately for studies with (1) high quality economic evaluations based

on the CHEC list, and (2) low risk of bias based on the Cochrane risk of bias tool. Two analyses (3 and 4) explored the impact of the value of the WTP per QALY (set at $40,000 in the main analysis), by repeating the analysis with WTP values of $20,000 and $80,000. In the last analysis (5) only studies were pooled for which the covariance could be calculated directly from author data.

Results

Figure 1 comprises the study selection flow. Table 1 and Table 2 present a detailed overview of all included studies and their outcomes. Additionally, an overview of all studies that were excluded after the full-text screening phase and the reason for exclusion can be found in the Appendix C.6. In total 6226 papers, conference abstracts and trial registrations were identified. Full texts were examined of 178 papers and finally data from 37 articles published between 1990 and March 2021 were extracted. From 200 relevant protocols, conference abstracts and trial registrations, follow up was needed for 93 records with 76 different corresponding authors. For those not needing follow up it was clear that data gathering was still ongoing or data was already published and included in the screening. The 76 corresponding authors were approached via e-mail to clarify whether data were already available to implement in our meta-analysis. One reminder was sent within two weeks if there was no response. This yielded four additional articles to include from three different authors. In total, 53 (69.7%) authors responded and for 50 of those responses (94.3%) cost-utility data were not (yet) available or authors were not willing to share data at this time.

Table 1. Characteristics of included studies.

Characteristic	N	References
Country		
United Kingdom	12	Bogosian et al. (2015) and Crombie et al. (2018), Deluca et al. (2020) [High risk], Deluca et al. (2020) [Low risk], Dixon et al. (2016), Duarte et al. (2017), Hollinghurst et al. (2010), Lovell et al. (2017), Morriss et al. (2019), Powell et al. (2020), Richards et al. (2020), and Wright et al. (2020)
Netherlands	8	Aardoom, Dingemans, van Ginkel, et al. (2016), Ferwerda et al. (2018), Geraedts et al. (2015), Gerhards et al. (2010), Kolovos et al. (2016), Lokman et al. (2017), Van Luenen et al. (2019), and Warmerdam et al. (2010)
Sweden	5	Holst et al. (2018), Jolstedt et al. (2018), Kraepelien et al. (2018), Lenhard et al. (2017), and Lindsäter et al. (2019)
Germany	4	Buntrock et al. (2017), Kählke et al. (2019), Nobis et al. (2018), and Röhr et al. (2021)
Australia	3	Dear et al. (2015), Moayeri et al. (2019), and Titov et al. (2015)
United States	2	Joesch et al. (2012) and Murphy et al. (2016)
Canada	1	Yan et al. (2019)
Italy	1	Hunter et al. (2017)
Spain	1	Romero-Sanchiz et al. (2017)
Targeted disorder		

Depression	16	Buntrock et al. (2017), Dixon et al. (2016), Duarte et al. (2017), Geraedts et al. (2015), Gerhards et al. (2010), Hollinghurst et al. (2010), Holst et al. (2018), Kolovos et al. (2016), Kraepelien et al. (2018), Nobis et al. (2018), Romero-Sanchiz et al. (2017), Titov et al. (2015), Van Luenen et al. (2019), Warmerdam et al. (2010), Wright et al. (2020), and Yan et al. (2019)
Anxiey	7	Dear et al. (2015), Joesch et al. (2012), Jolstedt et al. (2018), Kählke et al. (2019), Lindsäter et al. (2019), Morriss et al. (2019), and Powell et al. (2020)
Substance abuse	5	Crombie et al. (2018), Deluca et al. (2020) [High risk], Deluca et al. (2020) [Low risk], Hunter et al. (2017) and Murphy et al. (2016)
Depression / anxiety	5	Bogosian et al. (2015), Ferwerda et al. (2018), Lokman et al. (2017), Moayeri et al. (2019), and Richards et al. (2020)
Obsessive compulsive disorder	2	Lenhard et al. (2017) and Lovell et al. (2017)
PTSD	1	Röhr et al. (2021)
Eating disorders	1	Aardoom, Dingemans, van Ginkel, et al. (2016)
Method of diagnosis for inclusion		
Formal diagnosis	13	Buntrock et al. (2017) and Dear et al. (2015)[a], Dixon et al. (2016)[a], Hollinghurst et al. (2010), Holst et al. (2018)[a], Joesch et al. (2012), Jolstedt et al. (2018), Kolovos et al. (2016), and Lenhard et al. (2017)[a], Lindsäter et al. (2019), Lovell et al. (2017), Morriss et al. (2019), and Romero-Sanchiz et al. (2017)
Self-reported symptoms	23	Aardoom, Dingemans, van Ginkel, et al. (2016), Bogosian et al. (2015), and Crombie et al. (2018), Deluca et al. (2020) [High risk], Deluca et al. (2020) [Low risk], Duarte et al. (2017), Ferwerda et al. (2018), Geraedts et al. (2015), Gerhards et al. (2010), Hunter et al. (2017), Kählke et al. (2019), Kraepelien et al. (2018), Lokman et al. (2017), Moayeri et al. (2019), Murphy et al. (2016), Nobis et al. (2018), Powell et al. (2020), and Richards et al. (2020)[a], Röhr et al. (2021) and Titov et al. (2015)[a], Van Luenen et al. (2019), Warmerdam et al. (2010), and Wright et al. (2020)
Other	1	Yan et al. (2019) (no assessment before inclusion)
Recruitment type		
Via clinical institution	20	Deluca et al. (2020) [High risk], Deluca et al. (2020) [Low risk], Dixon et al. (2016), Duarte et al. (2017), Ferwerda et al. (2018), Hollinghurst et al. (2010), Holst et al. (2018), Hunter et al. (2017), Joesch et al. (2012), Kolovos et al. (2016), Kraepelien et al. (2018), Lovell et al. (2017), Moayeri et al. (2019), Morriss et al. (2019), Murphy et al. (2016), Richards et al. (2020), Romero-Sanchiz et al. (2017), Van Luenen et al. (2019), Wright et al. (2020), and Yan et al. (2019)
Open or mass media recruitment	13	Aardoom, Dingemans, van Ginkel, et al. (2016), Bogosian et al. (2015), Buntrock et al. (2017), Dear et al. (2015), Gerhards et al. (2010), Jolstedt et al. (2018), Kählke et al. (2019), Lindsäter et al. (2019), Nobis et al. (2018), Powell et al. (2020), Röhr et al. (2021), Titov et al. (2015), and Warmerdam et al. (2010)
Other[b]	4	Crombie et al. (2018), Geraedts et al. (2015), Lenhard et al. (2017), and Lokman et al. (2017)
Economic perspective		

Societal	22	Aardoom, Dingemans, van Ginkel, et al. (2016), Buntrock et al. (2017), and Crombie et al. (2018), Deluca et al. (2020) [High risk], Deluca et al. (2020) [Low risk], Dixon et al. (2016), Ferwerda et al. (2018), Geraedts et al. (2015), Gerhards et al. (2010), Holst et al. (2018), Jolstedt et al. (2018), Kählke et al. (2019), Kolovos et al. (2016), Kraepelien et al. (2018), Lenhard et al. (2017), Lindsäter et al. (2019), Lokman et al. (2017), Lovell et al. (2017), Nobis et al. (2018), Romero-Sanchiz et al. (2017), Van Luenen et al. (2019), and Warmerdam et al. (2010)
Healthcare	15	Bogosian et al. (2015), Dear et al. (2015), Duarte et al. (2017), Hollinghurst et al. (2010), Hunter et al. (2017), Joesch et al. (2012), Moayeri et al. (2019), Morriss et al. (2019), Murphy et al. (2016), Powell et al. (2020), Richards et al. (2020), Röhr et al. (2021), Titov et al. (2015), Wright et al. (2020), and Yan et al. (2019)
Follow-up period		
8-12 weeks	6	Dear et al. (2015), Jolstedt et al. (2018), Lenhard et al. (2017), Lindsäter et al. (2019), Moayeri et al. (2019), and Warmerdam et al. (2010)
4-6 months	6	Aardoom, Dingemans, van Ginkel, et al. (2016), Bogosian et al. (2015), Kählke et al. (2019), Nobis et al. (2018), Röhr et al. (2021), and Van Luenen et al. (2019)
8-9 months	2	Hollinghurst et al. (2010) and Murphy et al. (2016)
12-14 months	20	Buntrock et al. (2017) and Crombie et al. (2018), Deluca et al. (2020) [High risk], Deluca et al. (2020) [Low risk], Dixon et al. (2016), Geraedts et al. (2015), Gerhards et al. (2010), Holst et al. (2018), Hunter et al. (2017), Kolovos et al. (2016), Kraepelien et al. (2018), Lokman et al. (2017), Lovell et al. (2017), Morriss et al. (2019), Powell et al. (2020), Richards et al. (2020), Romero-Sanchiz et al. (2017), Titov et al. (2015), Wright et al. (2020), and Yan et al. (2019)
18 months or longer	3	Duarte et al. (2017), Ferwerda et al. (2018), and Joesch et al. (2012)
Intervention duration		
Shorter than 6 weeks	2	Kolovos et al. (2016) and Röhr et al. (2021)
6-8 weeks	15	Aardoom, Dingemans, van Ginkel, et al. (2016), Bogosian et al. (2015), Buntrock et al. (2017), Dear et al. (2015), Duarte et al. (2017), Geraedts et al. (2015), Gerhards et al. (2010), Kählke et al. (2019), Nobis et al. (2018), Powell et al. (2020), Richards et al. (2020), Titov et al. (2015), Van Luenen et al. (2019), Warmerdam et al. (2010), and Wright et al. (2020)
9-12 weeks	13	Crombie et al. (2018), Holst et al. (2018), Joesch et al. (2012), Jolstedt et al. (2018), Kraepelien et al. (2018), Lenhard et al. (2017), Lindsäter et al. (2019), Lovell et al. (2017), Moayeri et al. (2019), Morriss et al. (2019), Murphy et al. (2016), Romero-Sanchiz et al. (2017), and Yan et al. (2019)
Longer than 12 weeks	3	Dixon et al. (2016), Ferwerda et al. (2018), and Hollinghurst et al. (2010)
Undefined or unlimited access	4	Deluca et al. (2020) [High risk], Deluca et al. (2020) [Low risk], Hunter et al. (2017) and Lokman et al. (2017)
Human guidance available		

Yes	27	Aardoom, Dingemans, van Ginkel, et al. (2016), Bogosian et al. (2015), Buntrock et al. (2017), Dear et al. (2015), Dixon et al. (2016), Duarte et al. (2017), Ferwerda et al. (2018), Geraedts et al. (2015), Hollinghurst et al. (2010), Holst et al. (2018), Joesch et al. (2012), Jolstedt et al. (2018), Kählke et al. (2019), Kolovos et al. (2016), Kraepelien et al. (2018), Lenhard et al. (2017), Lindsäter et al. (2019), Lovell et al. (2017), Moayeri et al. (2019), Morriss et al. (2019), Murphy et al. (2016), Nobis et al. (2018), Richards et al. (2020), Titov et al. (2015), Van Luenen et al. (2019), Warmerdam et al. (2010), and Wright et al. (2020)
No	10	Crombie et al. (2018), Deluca et al. (2020) [High risk], Deluca et al. (2020) [Low risk], Gerhards et al. (2010), Hunter et al. (2017), Lokman et al. (2017), Powell et al. (2020), Röhr et al. (2021), Romero-Sanchiz et al. (2017), and Yan et al. (2019)
Guidance type		
Self-guided	10	Crombie et al. (2018), Deluca et al. (2020) [High risk], Deluca et al. (2020) [Low risk], Gerhards et al. (2010), Hunter et al. (2017), Lokman et al. (2017), Powell et al. (2020), Röhr et al. (2021), Romero-Sanchiz et al. (2017), and Yan et al. (2019)
Email or written support	11	Buntrock et al. (2017), Ferwerda et al. (2018), Geraedts et al. (2015), Jolstedt et al. (2018), Kählke et al. (2019), Kolovos et al. (2016), Kraepelien et al. (2018), Lindsäter et al. (2019), Nobis et al. (2018), Richards et al. (2020), and Warmerdam et al. (2010)
Chat support	1	Hollinghurst et al. (2010)
Telephone support	5	Dixon et al. (2016), Duarte et al. (2017), Lovell et al. (2017), Moayeri et al. (2019), and Van Luenen et al. (2019)
Video conferencing	1	Bogosian et al. (2015)
Face-to-face support	3	Joesch et al. (2012), Murphy et al. (2016), and Wright et al. (2020)
Combination	6	Aardoom, Dingemans, van Ginkel, et al. (2016), Dear et al. (2015), Holst et al. (2018), Lenhard et al. (2017), Morriss et al. (2019), and Titov et al. (2015)
Guidance frequency		
Not applicable	10	Crombie et al. (2018), Deluca et al. (2020) [High risk], Deluca et al. (2020) [Low risk], Gerhards et al. (2010), Hunter et al. (2017), Lokman et al. (2017), Powell et al. (2020), Röhr et al. (2021), Romero-Sanchiz et al. (2017), and Yan et al. (2019)
Less than weekly	3	Dixon et al. (2016), Hollinghurst et al. (2010), and Lovell et al. (2017)
weekly	21	Aardoom, Dingemans, van Ginkel, et al. (2016), Bogosian et al. (2015), Buntrock et al. (2017), Dear et al. (2015), Duarte et al. (2017), Ferwerda et al. (2018), Geraedts et al. (2015), Holst et al. (2018), Joesch et al. (2012), Jolstedt et al. (2018), Kählke et al. (2019), Kolovos et al. (2016), Lindsäter et al. (2019), Moayeri et al. (2019), Morriss et al. (2019), Nobis et al. (2018), Richards et al. (2020), Titov et al. (2015), Van Luenen et al. (2019), Warmerdam et al. (2010), and Wright et al. (2020)
More than weekly	3	Kraepelien et al. (2018), Lenhard et al. (2017), and Murphy et al. (2016)
Control condition type		

Care as usual[c]	32	Aardoom, Dingemans, van Ginkel, et al. (2016), Bogosian et al. (2015), Buntrock et al. (2017), and Dear et al. (2015), Deluca et al. (2020) [High risk], Deluca et al. (2020) [Low risk], Dixon et al. (2016), Duarte et al. (2017), Ferwerda et al. (2018), Geraedts et al. (2015), Gerhards et al. (2010), Hollinghurst et al. (2010), Holst et al. (2018), Hunter et al. (2017), Joesch et al. (2012), Kählke et al. (2019), Kolovos et al. (2016), Kraepelien et al. (2018), Lenhard et al. (2017), Lindsäter et al. (2019), Lokman et al. (2017), Lovell et al. (2017), Morriss et al. (2019), Murphy et al. (2016), Nobis et al. (2018), Powell et al. (2020), Richards et al. (2020), Röhr et al. (2021), Romero-Sanchiz et al. (2017), Titov et al. (2015), Warmerdam et al. (2010), and Yan et al. (2019)
Attention control	5	Crombie et al. (2018), Jolstedt et al. (2018), Moayeri et al. (2019), Van Luenen et al. (2019), and Wright et al. (2020)
Mode of delivery		
Website[d]	28	Aardoom, Dingemans, van Ginkel, et al. (2016), Buntrock et al. (2017), Dear et al. (2015), Duarte et al. (2017), Ferwerda et al. (2018), Geraedts et al. (2015), Gerhards et al. (2010), Holst et al. (2018), Hunter et al. (2017), Joesch et al. (2012), Jolstedt et al. (2018), Kählke et al. (2019), Kolovos et al. (2016), Kraepelien et al. (2018), Lenhard et al. (2017), Lindsäter et al. (2019), Lokman et al. (2017), Lovell et al. (2017), Murphy et al. (2016), Nobis et al. (2018), Powell et al. (2020), Richards et al. (2020), Romero-Sanchiz et al. (2017), Titov et al. (2015), Van Luenen et al. (2019), Warmerdam et al. (2010), Wright et al. (2020), and Yan et al. (2019)
Telephone or videoconferencing	4	Bogosian et al. (2015), Dixon et al. (2016), Moayeri et al. (2019), and Morriss et al. (2019)
Chat	1	Hollinghurst et al. (2010)
Text messaging	1	Crombie et al. (2018)
App	1	Röhr et al. (2021)
Game (app or website)	2	Deluca et al. (2020) [High risk], Deluca et al. (2020) [Low risk]

[a] These studies had inclusion criteria based on both (the absence of) a formal diagnosis and self-reported symptoms or a formal diagnosis was conducted after a self-reported symptom-based inclusion
[b] Could involve a mixture of recruitment strategies, targeting a specific group of people, or recruitment via companies
[c] Could involve a waitlist or do-nothing approach (where often participants were allowed to use other forms of treatment during the study period) or a one-time informational session or flyer
[d] Interventions consisted of (often weekly) modules with (cognitive-behavioral) exercises ($n = 27$) or web-based self-monitoring ($n = 1$)

Table 2. Sample and outcomes of included studies.

Author (year of publication)	Sample size	Mean age (SD)	% Female	Delta QALY (SE)	Delta costs in US dollars (SE)	INB in US dollars (VAR)
Aardoom, Dingemans, van Ginkel, et al. (2016)	354	24.2 (7.7)	99.0	$< .01(.01)$	-660 (433)	668 (489207)
Bogosian et al. (2015)	40	52.7 (9.5)	55.0	$-.01(.02)$	-3216 (2056)	2976 (4688465)

Buntrock et al. (2017)	406	45.0 (11.9)	73.9	$.01(.02)$	169 (179)	232 (884732)
Crombie et al. (2018)	825	35.0 (miss)	0.0	$-.01(.02)$	488 (357)	-728 (657104)
Dear et al. (2015)	70	65.5 (5.13)	60.0	$.01(.03)$	61 (19)	339 (1847789)
Deluca et al. (2020) [High risk]	756	16.1 (0.9)	50.2	$-.01(.01)$	547 (722)	-930 (737655)
Deluca et al. (2020) [Low risk]	883	15.2 (1.0)	51.7	$-.01(.01)$	639 (668)	-1058 (693923)
Dixon et al. (2016)	609	49.6 (12.8)	68.5	$< .01(.01)$	2805 (144)	-196 (281174)
Duarte et al. (2017)	691	39.9 (12.7)	67.0	$-.04(.04)$	171 (145)	-1911 (2622289)
Ferwerda et al. (2018)	133	56.4 (10.0)	64.9	$.06(.03)$	5035 (3112)	-2675 (10740997)
Geraedts et al. (2015)	231	43.4 (9.2)	62.3	$< .01(.03)$	-889 (2962)	889 (10926230)
Gerhards et al. (2010)	303	44.9 (11.6)	43.2	$-.01(.02)$	66 (1901)	-466 (4748077)
Hollinghurst et al. (2010)	297	34.9 (11.6)	68.0	$.03(.02)$	778 (106)	302 (730265)
Holst et al. (2018)	90	38.6 (11.7)	77.8	$-.05(.03)$	-41 (58)	-1718 (6932833)
Hunter et al. (2017)	763	Median 49 (IQR 35-61)	38.5	$< .01(< .01)$	3 (4)	21 (36931)
Joesch et al. (2012)	690	45.1 (13.2)	71.7	$.05(.04)$	257 (523)	1743 (3150182)
Jolstedt et al. (2018)	131	10.0 (1.3)	53.4	$< .01(.09)$	-347 (6)	42 (15261387)
Kählke et al. (2019)	264	43.3 (10.2)	73.1	$.01(.00)$	-374 (690)	670 (706242)
Kolovos et al. (2016)	269	38.0 (11.4)	53.9	$.03(0.15)$	571 (310)	138 (37849735)
Kraepelien et al. (2018)	945	43.0 (12.2)	72.9	$.01(.03)$	-46 (496)	-1519 (4559865)
Lenhard et al. (2017)	67	14.6 (1.7)	46.0	$< .01(< .01)$	19 (5)	121 (269039)
Lindsäter et al. (2019)	100	46.2 (8.8)	85.0	$< .01(.02)$	-71 (317)	243 (1068351)
Lokman et al. (2017)	220	44.2 (9.9)	59.1	$.02(.02)$	-5000 (2740)	5840 (9751347)
Lovell et al. (2017)	473	Median 33 (range 18-77)	60.3	$.01(.01)$	-25 (251)	465 (430364)

Moayeri et al. (2019)	110	68.1 (8.8)	65.5	$-.01(.01)$	-270 (23)	-54 (197204)
Morriss et al. (2019)	156	Median 32 (range 19-82)	69.2	$.07(.07)$	-1419 (1299)	4219 (23565882)
Murphy et al. (2016)	507	34.9 (10.9)	37.9	$< .01(.01)$	147 (200)	-187 (170067)
Nobis et al. (2018)	256	51.0 (12.0)	62.9	$.01(.01)$	1001 (999)	278 (1457706)
Powell et al. (2020)	2116	37.2 (13.8)	80.2	$.01(.01)$	-87 (82)	647 (51304)
Richards et al. (2020)	361	Median 29 (IQR = 18)	71.5	$.02(.01)$	125 (65)	707 (292998)
Röhr et al. (2021)	133	33.3 (11.2)	38.3	$< .01(.01)$	-124 (139)	-36 (64494)
Romero-Sanchiz et al. (2017)	296	42.9 (10.3)	75.7	$.08(.03)$	-265 (459)	3526 (1329167)
Titov et al. (2015)	54	65.3 (3.0)	70.4	$.01(< .01)$	33 (25)	447 (30236)
Van Luenen et al. (2019)	188	46.3 (10.6)	11.7	$.01(< .01)$	13 (171)	1180 (690364)
Warmerdam et al. (2010)	263	45.0 (12.1)	71.1	$.01(.01)$	276 (613)	124 (442031)
Wright et al. (2020)	139	15.0 (1.4)	64.0	$.03(.06)$	-20 (161)	1180 (5895741)
Yan et al. (2019)	1407	47.1 (17.0)	73.0	$.01(.01)$	-140 (36)	356 (1491)

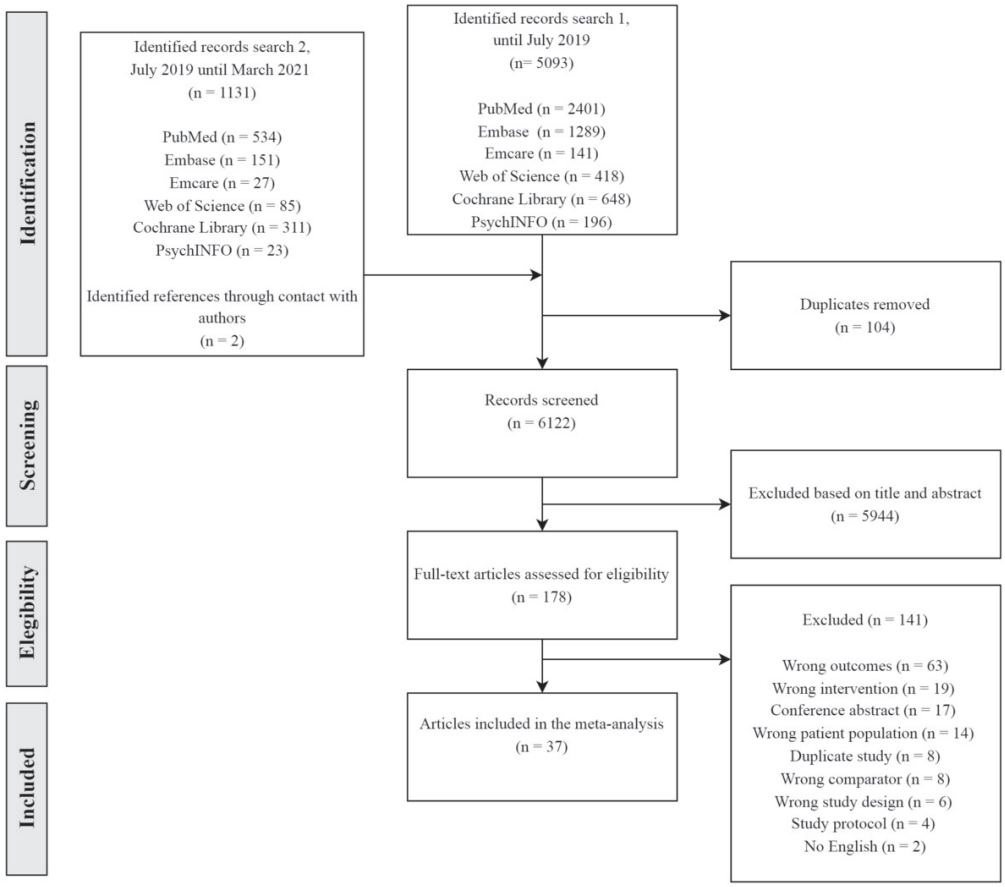
Miss=missing; IQR=Interquartile range; SD=Standard deviation; SE=standard error; VAR=variance.

Data preparation

For 7 studies no measure of spread of delta QALY or delta costs could, directly or indirectly, be deduced from the article. Data from one study were available within our research team (Aardoom, Dingemans, van Ginkel, et al., 2016). After inquiry with the authors of the other six studies (5 of which responded) data of only three comparisons were still missing. Two were solved by using Webplot Digitizer on the presented cost-effectiveness plane and the other by mean imputation based on two comparisons within the same study.

All authors of included studies were contacted to provide information on the covariance between delta QALY and delta costs. Authors responded for 31 (83.8%) of the 37 included studies, but not all authors were able to provide information on the covariance. Specifically, for 12 of the studies the covariance could be based on data from authors. The mean Pearson correlation between delta QALY and delta costs based on these 12 studies was $r = -.12$ ($SD = .16$). For the remaining 25 studies, covariances were calculated using Webplot Digitizer ($n = 13$) or using the estimated mean correlation calculated earlier ($n = 12$).

Figure 1. Study selection flow.



Characteristics of included studies

In total, the 37 included studies (Table 1) recruited 15,596 participants, ranging between 40 to 2,116. Some study conditions were irrelevant for this meta-analysis (e.g., an active intervention without internet component), so the main analysis was based on 14,946 participants. Intention-to-treat analyses were conducted in most studies ($n = 32$). Mental disorders that were targeted were depression ($n = 16$), anxiety ($n = 7$), alcohol or substance abuse ($n = 5$), depression and anxiety simultaneously ($n = 5$), obsessive compulsive disorder ($n = 2$), post-traumatic stress disorder ($n = 1$), and eating disorders ($n = 1$). Experimental interventions were mostly cognitive-behavioral based modules or websites that participants could engage with ($n = 28$). Other interventions consisted of teleconferencing ($n = 2$), text messaging ($n = 3$), a web-based game ($n = 2$) and telephone support ($n = 2$). In 27 studies some form of guidance within the intervention was available, whereas the intervention was self-guided in 10 studies. Guidance consisted of written feedback ($n = 11$), telephone calls ($n = 5$), face-to-face, including teleconferencing ($n = 4$), chat ($n = 1$), or a combination of these ($n = 6$). Control conditions of studies included waiting list or care as usual conditions ($n = 32$) and psychological placebo or attention control conditions ($n = 5$). Follow-up periods ranged from 8 weeks to 2 years, with most studies ($n = 20$) maintaining a 12 month follow-up period.

Quality of life

Questionnaires used to calculate QALYs were EQ-5D ($n = 32$), KIDSCREEN-10 ($n = 1$), SF-6D or SF-36 ($n = 3$) and the Australian quality of life instrument ($n = 1$). The pooled difference in effectiveness (intervention QALY gains minus control QALY gains) for included studies was .004 QALY ($SE = .002$), Hedges' $g = 0.05$ (95% CI 0.01; 0.10, $P = .016$). While the difference was statistically significant, likely because of the large sample size, the size of the difference was deemed negligible.

Costs

Main questionnaires used to measure healthcare use and costs in included studies were the Treatment Inventory of Costs in Psychiatric Patients ($n = 14$) and Client Service Receipt Inventory ($n = 7$), but other or self-made questionnaires, medical records and diaries were also used. In total, 15 studies reported costs from a healthcare perspective and 22 presented a societal perspective. The pooled difference in costs (intervention costs minus control costs) when studies with a healthcare and societal perspective were taken together was \$49 ($SE = 40$), Hedges' $g < 0.01$ (95% CI -0.10; 0.84, $P = .96$). Considering the uncertainty in measurements of costs, the small difference indicates that internet interventions were equally expensive as control conditions. Results for studies with different economic perspectives were similar, with no difference in costs for studies with a healthcare (\$40, $SE = 44$), Hedges' $g = -0.03$ (95% CI -0.21; 0.16, $P = .78$), and societal perspective (\$158, $SE = 76$), Hedges' $g = 0.03$ (95% CI -0.01; 0.10, $P = .15$).

Cost-effectiveness

Visual inspection of the individual INBs and their CIs (Figure 2) indicated substantial heterogeneity between the 37 included studies. Statistical measures of heterogeneity suggested otherwise, Cochran $Q(36) = 37.12$, $P = .42$, $I^2 = 3.0\%$ (95% CI 0.0%; 42.8%), but are difficult to interpret because of the large 95% CI of the I^2 and considerable within-study uncertainty. In other words, the between-study heterogeneity seemed to be overshadowed by the large within-study heterogeneity. Therefore, still a random-effects model was preferred over a fixed-effect model to pool INBs. The INB was positive (more favorable balance of costs and effects in the internet intervention compared to the control condition) in 25 of the included RCTs. Furthermore, at a WTP of $40,000 per QALY, the pooled INB was $255 (95% CI $91; $419, $P = .002$). The results suggest that internet interventions are slightly more cost-effectiveness compared to a do-nothing or care-as-usual approach.

Moderator analyses

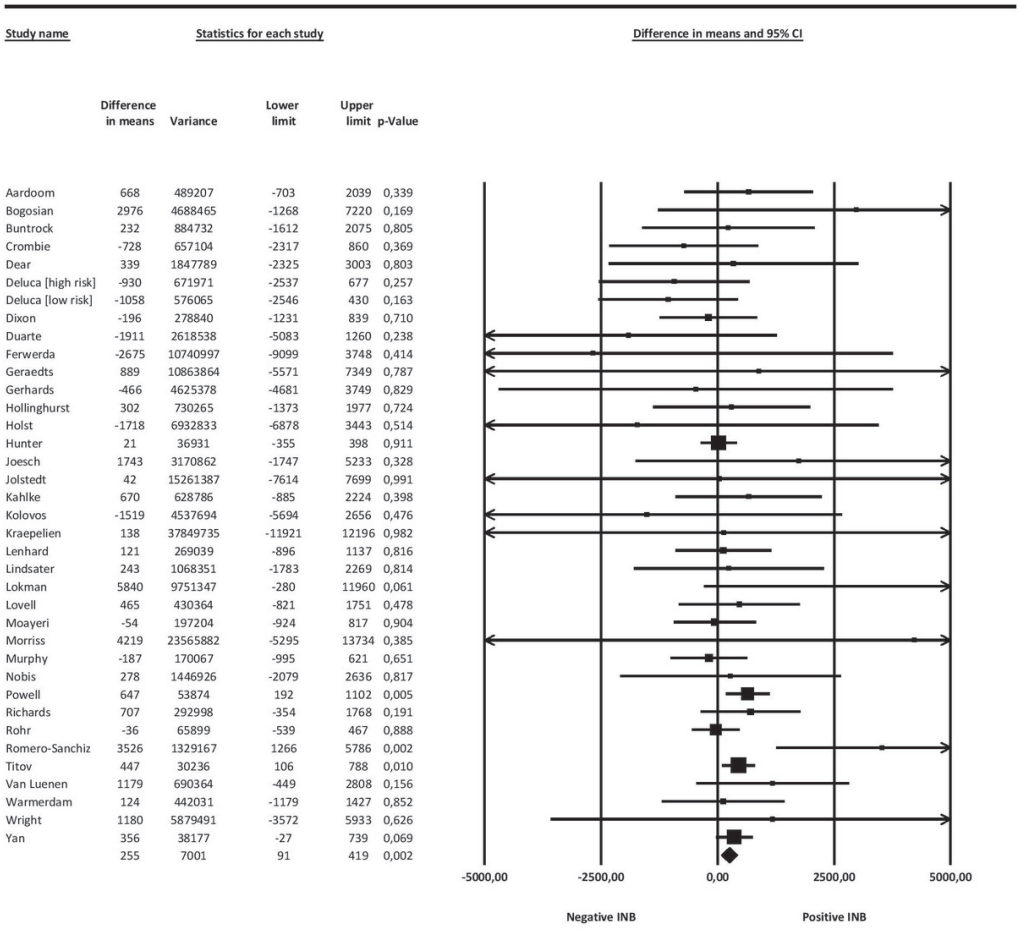
Pooled INBs were also calculated for subgroups based on the ten moderator variables. Outcomes for subgroups based on perspective, length of follow-up and targeted mental disorder are presented in the text and results for all moderator (including presence of guidance, intensity of guidance, type of guidance, recruitment strategy, diagnosis for inclusion, intervention duration, and control condition type) analyses can be found in the Appendix C.7.

Pooling studies based on economic perspective influenced the results. Specifically, looking at studies with a healthcare perspective separately ($n = 15$), the pooled INB was $280 (95% CI $109; $451, $P = .001$). For studies with a societal perspective ($n = 22$), the pooled INB was substantially lower at $161 (95% CI $-247; $569, $P = .44$). This suggests that cost-effectiveness of internet interventions compared to control conditions cannot be assumed when maintaining a societal perspective.

Studies with a short (shorter than 12 months) follow-up ($n = 14$) had a pooled INB of $112 (95% CI $-194; $418, $P = .47$) and studies with a long (12 months or longer) follow-up ($n = 23$) had a pooled INB of $270 (95% CI $-14; $554, $P = .063$). For RCTs with a long follow-up, statistical significance was likely not attained because the sample size decreased compared to the main analysis. Nevertheless, the studies with a long follow-up, which are usually better able to capture all relevant costs and effects than those with a short time horizon (Basu & Maciejewski, 2019; Neumann et al., 2016), had a pooled estimate comparable to the estimate of all studies taken together. This strengthens the idea that internet interventions are likely to be cost-effective compared to control conditions on the long term.

Concerning targeted mental disorders, a significant positive INB was found for internet interventions targeting anxiety ($n = 7$), $644 (95% CI $227; $1062, $P = .002$), and depression ($n = 16$), $387 (95% CI $156; $618, $P = .001$). Similarly, the INB of studies with internet interventions targeting depression and anxiety simultaneously (n=5), pooled INB of $580 (95% CI $-584; $1744, $P = .33$), and obsessive compulsive disorder (n=2), pooled INB of $253 (95% CI $-544; $1051, $P = .53$) was positive. The size of the INB for these two groups also indicated that cost-effectiveness compared to control conditions is likely, but statistical significance was not attained. Internet interventions were unlikely to be cost-effective

Figure 2. Forest plot of incremental net benefits (INB) and the pooled estimate according to a random effects model.



Note. Difference in means indicates the incremental net benefit. The last row indicates the pooled incremental net benefit according to a random effects model.

when targeting alcohol or substance abuse (n=5), pooled INB of $-129 (95% CI $-448; $191, $P = .43$). It must be noted that of the two studies regarding obsessive compulsive disorder one targeted children, further obscuring interpretability for this subgroup. Only one study was available for eating disorders and posttraumatic stress disorder, rendering it impossible to pool.

Risk of bias and quality of economic evaluation

Table 3 comprises details on risk of bias and Table 4 on quality ratings of the economic analysis for the included studies. The overall risk of bias of the included RCTs was considered high, with low risk of bias for 8, medium risk of bias for 8 and high risk of bias for 21 studies. The economic appraisal of the included studies, based on the CHEC list, suggested moderate quality, with 24 studies receiving a high and 13 a low quality rating. Level of agreement between raters was considered low for the risk of bias assessment (55% agreement) and high for the CHEC list ratings (89% agreement).

Publication bias

A funnel plot of included studies is presented in Figure 3. Visual inspection of the plot seemed to suggest some evidence for publication bias. Egger's test did not indicate asymmetry in the funnel plot ($z = -.026$, one-tailed $P = .80$).

Sensitivity analyses

Results of all sensitivity analyses can be found in the Appendix C.8. First, the main analysis was repeated for studies with a high quality rating on the CHEC list ($n = 24$), resulting in a pooled INB of $253 (95% CI $43; $463, $P = .018$). This suggests cost-effectiveness of internet interventions was maintained when looking at high quality studies alone and that results from the main analysis were not dependent on low quality studies. Secondly, the pooled INB for studies with a low risk of bias ($n = 8$) was $244 (95% CI $-555; $1042, $P = .55$). A similar result was found when medium risk of bias was considered as low risk of bias ($n = 16$), with a pooled INB of $216 (95% CI $-182; $615, $P = .29$). The pooled INB for RCTs with low risk of bias was comparable to that of the main analysis, suggesting that the pooled result of all 37 studies was not critically biased. However, for high risk of bias studies the pooled INB was slightly higher, $272 (95% CI $68; $475, $P = .009$). This suggests a small overestimation in the overall pooled estimate. Thirdly, when one QALY was valued at $20,000 instead of $40,000 the pooled INB of the 37 studies was $145 (95% CI $56; $234, $P = .001$). At a WTP of $80,000 for one QALY the pooled INB was $431 (95% CI $115; $747, $P = .008$). The two analyses show that if society's WTP for one additional QALY for an individual is lower ($20,000) or higher ($80,000) than $40,000, cost-effectiveness of internet interventions compared to care as usual is still expected. Finally, the pooled INB of studies for which the covariance could be calculated directly from author data ($n = 12$) was $264 (95% CI $-167; $694, $P = .23$). This was similar to the pooled estimate of all 37 studies, indicating that the way the covariance was calculated did not greatly influence the results of the meta-analysis.

Deviations from the protocol

An individual participant data meta-analysis was not achievable, so moderator analyses based on age, gender and symptom severity were not possible or feasible. Additionally, a moderator analysis with subgroups based on whether the study was conducted by developers of the interventions or not was planned, but this information proved too difficult to find for many of the studies.

Table 3. Cochrane risk of bias table of included studies.

Author (year of publication)	1	2	3	4	5	6	7	Overall Risk of Bias
Aardoom, Dingemans, van Ginkel, et al. (2016)	+	+	-	+	+	+	+	Low
Bogosian et al. (2015)	+	+	-	+	+	?	+	Low
Buntrock et al. (2017)	+	+	-	+	+	+	+	Low
Crombie et al. (2018)	+	+	+	+	+	+	+	Low
Dear et al. (2015)	+	+	-	-	?	?	+	High
Deluca et al. (2020) [High risk]	+	+	-	+	-	+	+	High
Deluca et al. (2020) [Low risk]	+	+	-	+	-	+	+	High
Dixon et al. (2016)	+	+	-	+	+	+	?	Medium
Duarte et al. (2017)	+	+	-	+	?	+	-	High
Ferwerda et al. (2018)	+	+	-	+	-	?	+	High
Geraedts et al. (2015)	+	+	-	+	+	?	+	Low
Gerhards et al. (2010)	+	+	-	+	-	?	+	High
Hollinghurst et al. (2010)	+	+	-	?	-	?	-	High
Holst et al. (2018)	+	+	-	-	?	+	?	High
Hunter et al. (2017)	+	+	-	-	+	+	-	High
Joesch et al. (2012)	+	+	-	+	-	+	+	High
Jolstedt et al. (2018)	+	+	-	+	+	+	+	Low
Kählke et al. (2019)	+	+	-	?	+	+	+	Medium
Kolovos et al. (2016)	+	?	-	+	+	+	+	Medium
Kraepelien et al. (2018)	+	+	-	+	+	?	+	Low
Lenhard et al. (2017)	+	+	-	+	+	+	-	High
Lindsäter et al. (2019)	+	+	-	+	+	+	+	Low
Lokman et al. (2017)	?	+	-	+	-	?	-	High
Lovell et al. (2017)	+	+	-	?	+	+	-	High
Moayeri et al. (2019)	+	+	-	+	+	?	?	Medium
Morriss et al. (2019)	+	+	-	+	+	?	?	Medium
Murphy et al. (2016)	+	+	-	?	+	?	-	High
Nobis et al. (2018)	+	+	-	+	+	-	+	High
Powell et al. (2020)	+	+	-	+	-	?	+	High
Richards et al. (2020)	+	-	-	-	-	+	+	High
Röhr et al. (2021)	+	+	-	?	+	+	+	Medium
Romero-Sanchiz et al. (2017)	+	+	-	+	+	?	?	Medium
Titov et al. (2015)	+	+	-	-	?	?	?	High

Van Luenen et al. (2019)	+	+	-	?	+	?	+	Medium
Warmerdam et al. (2010)	+	?	-	+	+	?	-	High
Wright et al. (2020)	+	+	-	+	-	?	?	High
Yan et al. (2019)	-	-	-	?	-	+	-	High

1=Random sequence generation, 2=Allocation concealment, 3=Blinding of participants and personnel, 4=Incomplete outcome data, 5=Blinding of outcome assessors, 6=Selective reporting, 7=Other bias
+=item scored as low risk of bias.
-=item scored as high risk of bias.
?=item scored as unclear risk of bias.

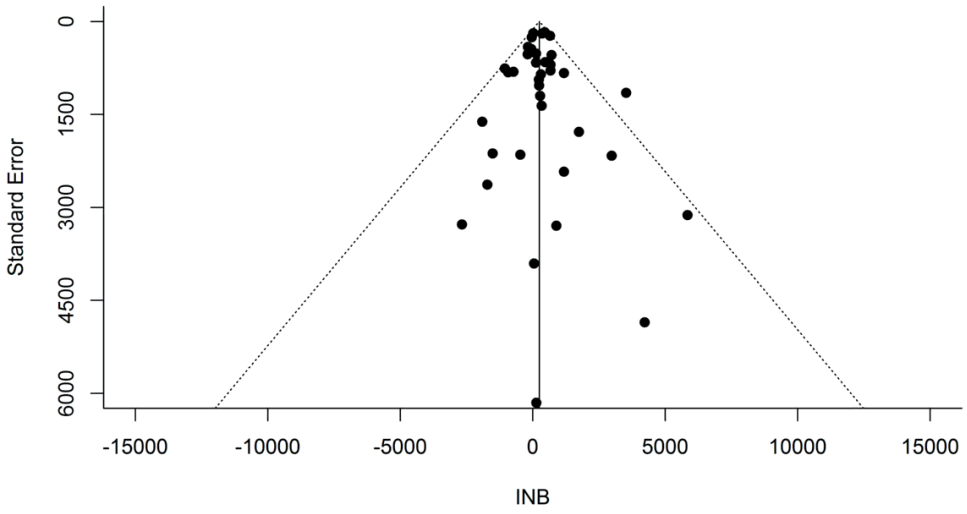
Table 4. Quality of the economic evaluation of included studies using the CHEC list.

Author (year of publication)	1	2	3	4	5	6	7[a]	8[a]	9[a]	10[a]	11[a]	12[a]	13[a]	14[a]	15	16	17	18	19	Sum score	Quality
Aardoom, Dingemans, van Ginkel, et al. (2016)	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	17/19	H
Bogosian et al. (2015)	+	+	+	+	-	-	+	+	+	+	+	+	+	+	-	+	-	+	-	14/19	H
Buntrock et al. (2017)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	19/19	H
Crombie et al. (2018)	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	18/19	L
Dear et al. (2015)	+	+	+	+	-	-	-	+	+	+	+	+	+	+	-	+	+	+	+	14/19	L
Deluca et al. (2020) [High risk]	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	+	-	15/19	L
Deluca et al. (2020) [Low risk]	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	+	-	15/19	L
Dixon et al. (2016)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	17/19	H
Duarte et al. (2017)	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	-	17/19	H
Ferwerda et al. (2018)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	17/19	H
Geraedts et al. (2015)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	18/19	H
Gerhards et al. (2010)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	18/19	H
Hollinghurst et al. (2010)	+	+	+	+	-	-	+	+	+	+	+	+	+	+	+	+	-	+	+	16/19	H
Holst et al. (2018)	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	-	+	-	16/19	L
Hunter et al. (2017)	+	+	+	+	-	-	+	+	+	+	+	+	+	+	+	+	-	+	-	15/19	H
Joesch et al. (2012)	+	+	-	+	+	-	-	+	+	+	+	+	+	-	-	+	+	+	-	13/19	L
Jolstedt et al. (2018)	+	+	+	+	-	+	+	+	+	+	+	+	+	+	-	+	+	+	+	17/19	H
Kählke et al. (2019)	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	18/19	H
Kolovos et al. (2016)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	17/19	H
Kraepelien et al. (2018)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	18/19	H
Lenhard et al. (2017)	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	18/19	H

Lindsäter et al. (2019)	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	-	17/19	H
Lokman et al. (2017)	+	+	-	+	+	+	+	+	+	+	+	-	-	+	+	-	+	+	-	14/19	L
Lovell et al. (2017)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	17/19	H
Moayeri et al. (2019)	+	+	+	+	-	-	+	+	+	+	+	+	+	+	+	+	+	+	-	16/19	H
Morriss et al. (2019)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	-	17/19	H
Murphy et al. (2016)	+	+	+	+	-	+	-	+	+	+	+	+	+	+	+	+	+	+	-	16/19	L
Nobis et al. (2018)	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	-	-	16/19	H
Powell et al. (2020)	+	+	-	+	+	+	+	-	+	+	+	+	+	+	-	+	+	+	-	14/19	L
Richards et al. (2020)	+	+	+	-	+	-	+	+	+	+	-	+	+	+	+	+	+	+	-	15/19	L
Röhr et al. (2021)	+	+	+	+	-	-	+	+	+	+	+	+	-	+	+	+	+	+	-	15/19	L
Romero-Sanchiz et al. (2017)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	-	17/19	H
Titov et al. (2015)	+	+	+	+	-	-	+	+	+	+	+	+	+	+	-	+	+	+	-	15/19	H
Van Luenen et al. (2019)	+	+	+	+	-	+	+	+	+	+	+	-	+	+	+	+	+	+	-	16/19	L
Warmerdam et al. (2010)	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	17/19	H
Wright et al. (2020)	+	+	-	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	-	16/19	H
Yan et al. (2019)	+	+	+	-	+	-	+	+	-	+	-	+	+	+	+	+	-	+	-	13/19	L

H=High overall quality; L=Low overall quality.
1=Is the study population clearly described?
2=Are competing alternatives clearly described?
3=Is a well-defined research question posed in answerable form?
4=Is the economic study design appropriate to the stated objective?
5=Is the chosen time horizon appropriate to include relevant costs and consequences?
6=Is the actual perspective chosen appropriate?
7=Are all important and relevant costs for each alternative identified?
8=Are all costs measured appropriately in physical units?
9=Are costs valued appropriately?
10=Are all important and relevant outcomes for each alternative identified?
11=Are all outcomes measured appropriately?
12=Are outcomes valued appropriately?
13=Is an incremental analysis of costs and outcomes of alternatives performed?
14=Are all future costs and outcomes discounted appropriately?
15=Are all important variables, whose values are uncertain, appropriately subjected to sensitivity analysis?
16=Do the conclusions follow from the data reported?
17=Does the study discuss the generalizability of the results to other settings and patient/client groups?
18=Does the article indicate that there is no potential conflict of interest of study researcher(s) and funder(s)?
19=Are ethical and distributional issues discussed appropriately?
[a] necessary item for a high quality score.
+=item scored as sufficient/high quality.
-=item scored as insufficient/low quality.

Figure 3. Funnel plot of standard error by incremental net benefit (INB) for inspecting publication bias.



Discussion

This systematic review and meta-analysis aimed to research the pooled evidence of the cost-effectiveness of internet interventions for mental disorders compared to control conditions. Results indicated that internet interventions were negligibly more effective in terms of QALY gains and equally costly compared to control conditions, but might be cost-effective. Internet interventions had an INB of $255 (95% CI $91; $419) compared to control conditions when society is willing to pay $40,000 for one QALY improvement for an individual. INBs were still positive when WTP for one additional QALY was lower ($20,000) or higher ($80,000). The results suggest that internet interventions for mental disorders are likely to be cost-effective compared to a do-nothing approach, especially when they target depression or anxiety and when some form of guidance is added to the intervention. This is especially interesting considering that internet interventions are now frequently and successfully being implemented (Titov et al., 2018) and can be used to serve populations that need mental health care but do not yet receive it (Ebert et al., 2018). Financing such scalable and, in some cases, anonymous interventions comes with difficulties, as they often do not fit in the funding possibilities for traditional treatments. Nevertheless, findings from this meta-analysis indicate that doing so might ultimately help to reduce the global burden of mental disorders.

Moderator analyses revealed that cost-effectiveness of internet interventions compared to control conditions was maintained for studies with a health care perspective, but not for those with a societal perspective. An explanation could be that indirect costs included in a societal perspective are usually higher and measured with more uncertainty than direct health care costs. Indeed, studies with a health care perspective included in the analyses had both a smaller cost range and smaller pooled standard error compared to studies with a societal perspective. A relatively small difference in health care costs might then be obscured by large indirect costs. This reflects a larger problem that sample size calculations for RCTs are almost exclusively based on detecting differences in disorder-specific effectiveness while neglecting QALYs and costs, rendering the trial unlikely to be adequately powered for an economic evaluation (Hollingworth et al., 2013). Furthermore, compared to control conditions, internet interventions were likely to be cost-effective for symptoms of anxiety, depression and obsessive compulsive disorder. This was not the case for alcohol and substance abuse. The finding that internet interventions for alcohol and substance abuse were not found to be cost-effective compared to controls, might be due to a lack of power in that subgroup. Alternatively, it could relate to the absence of guidance. Indeed, guided interventions were found to be cost-effective compared to care as usual whereas self-guided ones were not, and none of the alcohol and substance abuse studies incorporated guidance. Adding guidance to an internet intervention, then, might have a positive effect on cost-effectiveness. However, self-guided interventions have less functions than guided ones and do not compensate for this by higher levels of technical sophistication (i.e., responsiveness), at least for those targeting depression (Burger et al., 2020), perhaps partly explaining the positive influence of guidance in internet interventions. Another result was that studies with a follow-up period of 12 months or longer generally showed internet interventions for mental disorders to be efficient compared to a control condition, while this was not the case for studies with shorter follow-up periods. Possibly, some factors that influence the efficiency of internet interventions become apparent after a certain time only. For example, effects of an intervention concerning health

care visits or work productivity may take a while to attain. This substantiates the idea that follow-up periods of at least 12 months are important when conducting an economic evaluation of such interventions (Basu & Maciejewski, 2019; Neumann et al., 2016). Several other findings of the subgroup analyses are worth mentioning. First, recruiting participants through (social) media rather than by clinical referral was more likely to yield a positive INB. Speculatively, studies that used an open recruitment system included participants that were less severely ill or more strongly motivated for this type of interventions, hence benefitting more. Comparing study subgroups based on severity of symptoms, motivation to change, or related participant characteristics directly was not feasible for this meta-analysis, but could be an interesting avenue for future research. Second, shorter interventions were likely to be efficient compared to controls, whereas this was not true for longer interventions. It might be that longer internet interventions cost more, but do not perform better compared to shorter interventions in terms of QALYs or health care or societal costs. Third, cost-effectiveness of internet interventions for mental disorders was not likely when control conditions included an active component (i.e., attention control). Accordingly, activating participants seems to be an important ingredient for an efficient intervention, while the content may be of lesser importance. Moderator analyses should be considered as exploratory, because of the low number of studies in some subgroups.

QALYs and mental health interventions

Interestingly, internet interventions were found to be only marginally more effective than control conditions in terms of QALY gain. It is surprising that internet interventions produced practically the same amount of QALYs as a do-nothing approach, given the substantial evidence for the effectiveness of such interventions (Andersson et al., 2019). However, it is likely that the internet interventions were more effective in terms of symptom reduction and other areas of well-being, but these improvements were not captured well by the generic health-related quality-of-life measures (e.g., EQ-5D and SF-36) used in economic evaluations. Such instruments capture only a selective amount of domains of quality of life and employ an almost exclusive focus on people's current functional abilities with little emphasis on coping capabilities and resources (Pietersma et al., 2013). Consequently, they might not be suitable in contexts outside of health, such as chronic illness, elderly care, general well-being and mental illness (Mitchell et al., 2017). Future economic analyses of internet interventions for mental disorders should consider using other instruments, such as the ICECAP-A (Al-Janabi et al., 2012), which measures well-being beyond (physical) health, to complement generic health questionnaires (Keeley et al., 2016; National Institute for Health and Care Excellence, 2016).

Meta-analyses on cost-effectiveness data

To our knowledge, meta-analyses on cost-effectiveness studies have only been attempted in five studies in different fields of medicine by one other research team (Bagepally et al., 2020; Bagepally et al., 2019; Chaiyakittisopon et al., 2021; Haider et al., 2019; Noparatayaporn et al., 2021), but not yet in the area of mental health interventions. The theoretical method by Crespo et al. (2014) has been practically applied and explained by Bagepally et al. (2020).

The current study built on this method by improving the precision of estimating covariances between delta costs and delta QALYs. In general, many data are necessary to pool cost-effectiveness outcomes in a meta-analysis, some of which are often not reported. Enhancing quality of economic evaluations by clearly reporting costs, QALYs and indicators of spread (e.g., SD or SE) for all studied conditions or open availability of study data makes conducting meta-analyses on cost-effectiveness data more feasible. Besides these practical challenges, the unavoidable and substantial heterogeneity between cost-effectiveness studies alongside RCTs might make statistically pooling outcomes undesirable (Shields & Elvidge, 2020). Therefore, careful planning and cautious interpretation of the findings are warranted when considering a meta-analysis on cost-effectiveness data. Nevertheless, bearing the limitations in mind, several considerations corroborated the choice for pooling cost-effectiveness data. First, by adhering to clear predefined inclusion and exclusion criteria throughout the search procedure, the included studies were comparable across many characteristics. For example, the interventions were often similar in terms of content, mode of delivery, duration, and frequency. Additionally, in the majority of studies the control conditions consisted of a do-nothing approach (i.e., participants did not receive the internet intervention, but were allowed to keep on receiving the care they got before the study). Relatedly, 21 of the included studies used the TiC-P or CSRI to measure service use, suggesting that these studies considered the same costs, and almost all studies used the EQ-5D for calculating QALYs. Second, modeling studies were excluded to avoid additional variation between methods. Third, characteristics that were thought to have a large influence on the outcomes were tested in moderator and sensitivity analyses. Overall, these analyses were in line with the overall pooled estimate, further substantiating the robustness of the finding that internet interventions are likely to be cost-effective compared to control conditions. Lastly, the variance of the INB of individual studies was often large. Consequently, a meta-analysis was important, since it offered insight beyond individual studies, which are rarely powered to detect differences in QALYs and costs (Hollingworth et al., 2013).

Limitations

While the main, moderator and sensitivity analyses point in the direction of internet interventions being cost-effective compared to control conditions, findings should be interpreted with considerable caution and are difficult to generalize beyond the current sample of included studies. A first limitation is that the studies were heterogeneous in terms of included costs and follow-up duration, which warrants careful interpretation of the overall pooled result. Second, variances of the cost-effectiveness outcomes were large to such an extent that they overshadowed differences between studies, making it hard to understand and quantify between-study heterogeneity. Third, publication bias possibly shifted results in favor of internet interventions, because economic analyses are often a last step in effectiveness research and some results might never get published. For example, some contacted authors who published a protocol mentioning an economic analyses replied that such an analysis was ultimately not attempted, since the internet intervention was not found to be effective. To counter this problem, researchers are encouraged to designate in their RCT study protocol whether an economic analysis will be attempted and perform it regardless of results on effectiveness. A fourth reason to interpret the results with caution is that risk of bias in

most RCTs was high, possibly leading to a slight overestimation of the efficiency of internet interventions compared to control conditions. However, when quality was assessed with the CHEC list, which is arguably more suitable in the case of economic evaluations, high quality studies showed a positive INB whereas low quality studies did not. This confirms the idea that internet interventions might be cost-effective compared to controls only when the economic evaluation is of adequate quality. The subjectively chosen, though theory driven, cut-off points for the risk of bias and CHEC list should be taken into account when considering the importance of these sensitivity analyses. Indeed, studies designated as high quality based on the CHEC might not have had an appropriate time horizon or economic perspective.

Implications and future directions

The current study gives an overview of published articles on the cost-effectiveness of internet interventions for mental disorders and their findings and provides insights for future research and implementation steps. First, it must be noted that the results should be replicated in other meta-analyses before clinical and policy implications can be stated reliably. To accomplish this, standardization of economic evaluations in the area of mental health, as has been done in other sectors (Dirksen & Evers, 2016; Hiligsmann et al., 2019), would be helpful. Consequently, included studies in similar meta-analyses would be more homogeneous and easier to compare. Nevertheless, the pooled estimates of the main and subgroup analyses complement the results of the individual studies. They suggest that policy makers, insurance companies and subsidy providers should invest in internet interventions for mental disorders, as they appear to be an efficient way of improving quality of life compared to not offering them. For clinical practice, often facing financial pressure and waiting lists, this might involve a transition where internet interventions are increasingly embraced and become an integral part of the treatment options. The results of this meta-analysis also suggest that some components of internet interventions for mental disorders, such as recruitment strategy, symptom severity, guidance, and length of the intervention, are worth considering upon implementation. This is the first study to pool cost-effectiveness outcomes in the area of psychiatry, paving the way for other researchers to apply this method to new meta-analyses to further enhance the understanding of the cost-effectiveness of internet interventions compared to alternatives. For example, the current study only considered control conditions, as internet interventions are often used for early detection and underserved populations, for which care as usual and a waiting list are realistic comparators. Future work could compare internet interventions with active comparators such as face-to-face treatment, to clarify the difference in effects and costs between these forms of treatment. Additionally, this study looked broadly at the topic of cost-effectiveness of internet interventions for mental disorders and should be considered a starting point. It might be valuable to investigate a study sample with more homogeneous interventions or designs for more precise estimates of cost-effectiveness. Relatedly, modelling studies were not considered in this meta-analysis as pooling outcomes from such studies with those obtained from RCTs was undesirable. Performing a meta-analysis on modelling studies is feasible, however, and might be an interesting extension. Importantly, no study from a non-Western culture or low-income country was included in the study, while especially low-income countries or people in areas where health care is not paid for by the

government might benefit from internet interventions. Initiating internet interventions in such contexts might be a challenge, but can help to reduce health care costs in the long term. Conducting research on the cost-effectiveness of internet interventions in low-income countries is therefore highly commended.

Conclusion

Pooling outcomes of 37 studies revealed a small benefit of internet interventions for mental disorders compared to control conditions. Perspective of the economic evaluation, targeted mental disorder and WTP for one QALY moderated results. Generalizability to new studies is poor given the large variance of the outcome of interest and heterogeneity between studies. The findings show that cost-effectiveness of internet interventions for mental disorders compared to a do-nothing or care-as-usual approach is likely, but not guaranteed. Continuation of high quality and adequately powered economic evaluations is necessary. This is the first study in the area of psychiatry to pool cost-effectiveness outcomes in an aggregate-data meta-analysis, paving the way for other researchers to use and expand this method.