**Digital interventions for generalised anxiety disorder (GAD): systematic review and network meta-analysis**

**Saramago\* P1, Gega L2,3,4, Marshall D5, Nikolaidis FG6, Jankovic D1, Melton H**5**, Dawson S7, Churchill R5,7 and Bojke L1**

1 Centre for Health Economics, University of York

2 Department of Health Sciences, University of York

3 Hull York Medical School, University of York

4 Tees, Esk and Wear Valleys NHS Trust

5 Centre for Reviews & Dissemination, University of York

6 IQVIA

7 Common Mental Disorders Group, Cochrane Collaboration

\*Corresponding author: Pedro Saramago, Centre for Health Economics, Alcuin A Block, University of York, Heslington, York, YO10 5DD, UK; Tel: +44 (0)1904 321401; Fax: +44 (0)1904 321402

# Keywords

Worry, anxiety, cognitive behaviour therapy (CBT), mobile applications, digital, systematic review (SR), meta-analysis

#Words: 6,300

#Figures: 2

#Tables: 3

# ABSTRACT

**Background:** Generalised anxiety disorder is the most common mental health condition based on weekly prevalence. Digital interventions have been used as alternatives or as supplements to conventional therapies to improve access, patient choice and clinical outcomes. Little is known about their comparative effectiveness for generalised anxiety disorder.

**Methods:** We conducted a systematic review and network meta-analysis of randomised controlled trials comparing digital interventions with medication, non-digital interventions, non-therapeutic controls and no intervention.

**Results:** We included 21 randomised controlled trials with a total of 2,350 participants from generalised anxiety disorder populations. Pooled outcomes using analysis of Covariance and rankograms based on the surface under the cumulative ranking curves indicated that antidepressant medication and group therapy had a higher probability than digital interventions of being the ‘best’ intervention. Supported digital interventions were not necessarily “better” than unsupported (pure self-help) ones.

**Conclusions:** Due to very wide confidence intervals, network meta-analysis results were inconclusive as to whether digital interventions are better than no intervention and non-therapeutic active controls, or whether they confer an additional benefit to standard therapy. Future research needs to compare digital interventions with one-to-one therapy and with manualised non-digital self-help and to include antidepressant medication as a treatment comparator and effect modifier.

# INTRODUCTION

Generalised Anxiety Disorder (GAD) is the most common mental health condition with 6% point-prevalence (measured over the preceding week) in the UK, nearly double that of depression (3.3%) (Mental Health Foundation, 2016). It is often confused with panic disorder or depression when self-reported by survey participants ([McManus, Bebbington, Jenkins, & Brugha, 2016](#_ENREF_40)). GAD is characterised by excessive worry that persists for several months and leads to significant distress or impairment in everyday life and functioning ([World Health Organization, 2018](#_ENREF_68)). Other typical characteristics include free-floating anxiety and physical symptoms, such as muscle tension, headaches, restlessness, difficulty concentrating, irritability, or sleep problems. GAD is associated with low quality of life and high healthcare costs ([Revicki et al., 2012](#_ENREF_49)).

Psychological interventions can be effective for GAD, especially cognitive behaviour therapy (CBT) ([Hunot, Churchill, Teixeira, & de Lima, 2007](#_ENREF_32)) and applied relaxation ([Hayes-Skelton, Roemer, Orsillo, & Borkovec, 2013](#_ENREF_26)). CBT helps the individual challenge or tolerate worrying thoughts and confront anxiety-provoking situations rather than avoiding them. Applied relaxation counteracts the physical symptoms of GAD thought a series of tense-then-release muscle exercises that reduce muscle tension. Antidepressant medication can also be effective ([Slee et al., 2019](#_ENREF_55)) and is often the first choice for treatment by clinicians in view of limited capacity to deliver psychological interventions. To improve access to psychological therapies and increase patient choice and therapist capacity, digital interventions have been used as alternatives or supplements to conventional face-to-face clinic-based therapy (Aboujaoude, Gega, Parish, & Hilty, 2020; Gega & Gilbody, 2015).

Digital interventions are defined as software-based therapeutic activities accessed via technology platforms, such as the internet, virtual reality (VR), mobile phones. According to the World Health Organization (WHO), the term ‘digital intervention’ represents a discrete function of using technology to achieve health sector objectives (WHO, 2019). In the context of a specific condition, such as GAD, digital intervention (DIs) fulfil the discrete function of using software and digital media to deliver therapeutic activities that aim to improve symptoms associated with the condition in populations.

An example of a DI for GAD is a 10-week internet-based self-help programme consisting of psychoeducation (information about worry, stress and anxiety, including its risk factors and treatments), CBT (dealing with the purpose, meaning and content of worry, as well as modifying unhelpful responses to worry), relaxation (tensing then releasing body muscle groups and refocus attention away from worry) and physical activity (Christensen 2014a). Another example is a mobile app (Pham et al 2016) that teaches diaphragmatic breathing in a series of mini-games, from sailing a boat down a river to flying balloons into the sky.

Previous reviews of the effectiveness of DIs for GAD ([Cuijpers et al., 2014](#_ENREF_14); [Richards, Richardson, Timulak, & McElvaney, 2015](#_ENREF_50)) included mixed populations of anxiety disorders and depression without reporting outcomes separately for GAD subgroups within these mixed samples. Reporting a disorder-specific outcome for mixed samples can be misleading because it implies that, if an intervention works for the mixed sample, it will also work for each of its constituent populations. Studies reporting findings from mixed samples do not answer the question of whether DIs are effective for GAD to inform disorder-specific clinical guidelines. To achieve this, and while preserving the benefits of randomisation, we need to analyse GAD outcomes reported separately for GAD populations and GAD sub-samples within mixed populations.

This paper reports a systematic review and quantitative synthesis of RCTs comparing DIs with other interventions, non-therapeutic control arms and no intervention for GAD populations with varying levels of illness severity (sub-threshold, mild, moderate, severe). The review had four objectives:

1. Categorise the DIs and comparator arms into groups that could be pooled together.
2. Compare the pooled outcomes of DIs with the pooled outcomes of non-digital interventions, medication, non-therapeutic controls and no intervention for GAD symptoms.
3. Compare the pooled outcomes of different types of DIs.
4. Identify limitations and gaps in the existing research on DIs for GAD.

After describing the review, the classification and synthesis methodology in Section 2, Section 3 of the paper goes through the characteristics of the evidence base and synthesis results. Section 4 discusses the findings, followed by concluding remarks in Section 5.

# METHODS

**2.1. Protocol**

The protocol for the review was registered with PROSPERO 2018 CRD42018105837 as part of a larger piece of work that investigated the costs and outcomes of digital interventions for mental health, funded by the UK’s National Institute for Health Research (NIHR). The review has been conducted and reported as recommended by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension statement for network meta-analysis ([Moher, Liberati, Tetzlaff, & Altman, 2009](#_ENREF_42)).

**2.2. Search Strategy**

In December 2018, the following databases were searched to identify published and unpublished studies: MEDLINE, PsycINFO, Cochrane Central Register of Controlled Trials (CENTRAL), Cochrane Database of Systematic Reviews (CDSR), Cumulative Index to Nursing & Allied Health (CINAHL Plus), Database of Abstracts of Reviews of Effects (DARE), EMBASE, Web of Science Core Collection, NHS Economic Evaluation Database (NHS EED); Database of promoting health effectiveness reviews (DoPHER) and Proquest.

We also searched two clinical trial registries and other resources for ongoing studies: ClinicalTrials.gov and the WHO International Clinical Trials Registry Platform portal, as well as the NIHR portfolio of studies. Web searches were conducted using Google and Google Scholar making use of simplified search terms. After searches were complete, supplementary searches were conducted, including reference lists of included studies and forward citation searches. Finally, we contacted the authors of included studies for information on any other work in the field they were aware of. The searches were undertaken for studies conducted since 1997 and restricted to those written in English.

In June 2019, the searches were updated and widened to include terms based on unspecified Anxiety Disorders. An additional pilot search was conducted on Cochrane Library and PsycINFO databases using terms based on “worry” and “anxiety prevention”. This ensured that no articles were missed. No new included articles emerged from the pilot search, and it was not deemed necessary to expand it to all remaining databases. The full search terms and outputs of the database searches are provided in Appendices A (December 2018 search) and B (June 2019 search).

**2.3. Study Identification and Selection**

Two reviewers (DM, HM) independently screened all titles and abstracts of the identified studies against our inclusion/exclusion criteria. If either reviewer indicated a study could be relevant, we retrieved the full text. The same two reviewers independently assessed the full texts against our inclusion/exclusion criteria. A third reviewer (LG) resolved any disagreements through discussion to agree a final list of included and excluded studies.

Eligible studies included: a) study design: RCTs, to minimise risk of bias and confounding variables; b) participants: participants with symptoms or risk of GAD within mental health populations or within the general population; we defined this as a certified diagnosis using a standardised diagnostic interview or a score above an accepted cut-off for diagnosable GAD (which may include sub-threshold scores) in standardised questionnaires; c) interventions: software-based systems and technology platforms designed for patient-facing delivery of a mental health intervention (i.e. an intervention to improve mental health outcomes); d) comparisons: all comparisons relevant to DIs, even when two or more DIs were compared with each other without other comparators; and e) outcomes: GAD-specific measures of anxiety or worry (e.g. GAD-7), reported for GAD populations or GAD sub-samples within mixed populations.

We have excluded: i) mixed populations of GAD with other conditions, when the outcomes were not reported separately for GAD subgroups; ii) technology used as a means for telecommunication (e.g. email, phone or video) without any software-based processing; iii) software-based systems designed for training of health professionals or for administration without any patient-facing intervention components; and iv) studies that were only identified as protocols, abstracts or reviews; these were marked so we could check for RCTs that we may have missed in the database searches.

**2.4. Data extraction and risk of bias assessment**

Two researchers (DM, HM) independently extracted data from published and unpublished study reports. Data were extracted on the sample, study design, intervention and comparator characteristics, baseline characteristics, and results. Any discrepancies were resolved by a third reviewer (LG). Risk of bias of each study was assessed using Cochrane's Risk of Bias (RoB) 2.0 ([Sterne et al., 2019](#_ENREF_58)).

**2.5. Classification of DIs and their alternatives**

In order to conduct an evidence synthesis, it was necessary to classify and group the interventions from studies. We developed a classification system for DIs and their comparator interventions and controls, in the following four steps:

1. We conducted a detailed data extraction of intervention/control arm characteristics as reported within our included RCTs for GAD and their relevant linked papers.
2. We identified common and differentiating features of intervention/control arms between RCTs, but also incremental differences between interventions/controls within the same RCT.
3. We consulted the literature and an advisory group of health services researchers and clinicians about intervention features that could be important for clinical outcomes (e.g. amount of interpersonal contact, who offers support to the DI, whether the intervention is available publicly or via referral to specialist services, types of software required or therapeutic approaches used).
4. We applied the classification criteria to each randomisation arm in the included RCTs so that each intervention/control arm was assigned to a classification group.

These 4 steps were iterative; we resolved discrepancies by refining our classification criteria until two reviewers (DM, LG) independently reached the same allocation for every intervention/control they classified. We grouped DIs and their alternatives according to the three criteria below.

*Criterion 1 - Intervention (I) or Control (C):* An Intervention (I) was an action carried out as part of a research protocol for therapeutic purposes, i.e. it was expected to improve clinical symptoms and functioning based on psychological or behavioural theories and preliminary evidence. A Control (C) was a non-therapeutic activity that was not expected to make a clinical difference to the condition; this could be a psychological placebo, an ‘attention control’, or a change in usual care introduced by the research team to keep participants safe and minimise attrition.

*Criterion 2 - Digital (D) or Non-Digital (NoD)*: A Digital Intervention (DI) or a Digital Control (DC) included software programmes to guide patient-facing activities. A Non-Digital Intervention / Non-Digital Control (NoDI / NoDC) did not involve any technology and was delivered by printed materials or during face-to-face meetings, or via telecommunications technology without automated software e.g. consultations by email, skype, or phone.

*Criterion 3 - Supported (S) or Unsupported (U)*: Supported interventions/controls included scheduled or regular two-way person-to-person contact (e.g. between service user and clinician or researcher, or peer-to-peer). Unsupported interventions/controls either had no interpersonal contact or included limited ad-hoc interaction (e.g. phoning a helpline with any problems as a one-off). We also classified as unsupported interventions/controls those in which communication was one-way, such as a reminders by email, post or phone.

Based on these three criteria, we mapped DIs and their alternatives into eight groups resulting from the combinations of (I or C) x (D or NoD) x (S or U).

*Group 1: Supported Digital Intervention (SDI)*, e.g. computerised cognitive behaviour therapy with phone support; clinician-delivered therapy assisted by virtual reality.

*Group 2: Unsupported Digital Intervention (UDI)*, e.g. internet self-help without any clinician contact, mobile app with automated reminders but without personal interaction.

*Group 3: Supported Non-Digital Intervention (SNoDI)*, e.g. individual or group therapy in a clinic, or therapy delivered by a clinician via phone or an online platform.

*Group 4: Unsupported Non-Digital Intervention (UNoDI)*, e.g. self-help using a treatment manual or a book or a website without clinician input.

*Group 5: Supported Digital Control (SDC)*, e.g. access to a general health education website with weekly check-in calls from a researcher; virtual reality ‘placebo’ environment used in a clinic with a researcher present.

*Group 6: Unsupported Digital Control (UDC)*, e.g. access to an educational website without any support from a person or with just automated reminder emails.

*Group 7: Supported Non-Digital Control (SNoDC)*, e.g. weekly check-in by phone or regular clinical assessment face-to-face without any specific therapy instructions.

*Group 8: Unsupported Non-Digital Controls (UnoDC)*, e.g. printed materials with general health advice without any specific therapy instructions.

Separate groups were used for medication and no intervention. Medication (M) was any pharmacological agent (pills, injections, etc) offered as part of a research protocol. No Intervention (NI) included waiting lists and usual care in which there were no additional therapeutic activities and no changes in patient routines. The NI group may still have received medication or consultations as part of routine care, but this would have been equally accessible to all participants in a trial irrespective of group allocation, so the effect would be cancelled out across randomisation arms.

**2.6. Data synthesis and statistical analysis**

Over the last two decades, network meta-analysis (NMA) methods ([Lumley, 2002](#_ENREF_37)) also known as mixed treatment comparisons ([Caldwell, Ades, & Higgins, 2005](#_ENREF_9); [Lu & Ades, 2004](#_ENREF_36)) - have been developed to synthesise evidence from multiple studies. NMA is an extension to the standard (pairwise) meta-analysis, which pools together the results of studies for one type of intervention compared to one type of alternative (e.g. active treatment or placebo control) An NMA enables the simultaneous comparison of multiple interventions and multiple comparators within a single coherent analysis. Such an approach is routinely used in health technology assessments to inform the optimal intervention strategy for a given medical condition ([Cooper et al., 2011](#_ENREF_12)). NMAs are often used to inform estimates of clinical and cost-effectiveness and commissioning decisions.

In the NMA, an ANalysis of COVAriance (ANCOVA) modelling framework was used, where a final outcome measurement is synthesised and adjusted for baseline measurements. Compared to the ‘change from baseline’ approach, the ANCOVA model avoids guessing within-patient correlation across measurements as typically this is not reported in studies. Treatment effect estimates based on ANCOVA methods have been shown to be more efficient, less biased and robust to random baseline imbalance ([Dias, Sutton, Ades, & Welton, 2013](#_ENREF_19); [Dias, Welton, Sutton, & Ades, 2014](#_ENREF_20); [Higgins & Cochrane Collaboration, 2020](#_ENREF_28); [Riley et al., 2013](#_ENREF_52); [van Breukelen, 2013](#_ENREF_64); [Winkens, van Breukelen, Schouten, & Berger, 2007](#_ENREF_67)). Hence, the ANCOVA model, is the preferred method for estimating treatment effects from continuous outcomes ([Deeks, Higgins, & Altman, 2008](#_ENREF_18); [Fu et al., 2008](#_ENREF_23); [Vickers, 2001](#_ENREF_65); [Vickers & Altman, 2001](#_ENREF_66)).

We adopted a modelling approach in line with the parameterisation for continuous data with normal likelihood and identity link used by Dias et al ([Dias, Sutton, et al., 2013](#_ENREF_19); [Dias, Welton, Sutton, & Ades, 2014](#_ENREF_20)). Fixed-effects (FE) and random-effects (RE) models (the latter accounting for potential correlation within multi-arm trials) were fit to the data. In the model, patients who did not receive any treatment were assumed to neither improve nor worsen over the duration (i.e. null placebo effect). Furthermore, it was assumed that the effect of the baseline measurement is common across all treatments, implying that when two active treatments are compared in a trial, the baseline effects are offset.

All analyses were conducted within a Bayesian Markov chain Monte Carlo (MCMC) approach, fitted using WinBUGS software version 1.4.3 (Copyright © 2007 Medical Research Council (UK) and Imperial College (UK))([Lunn, Thomas, Best, & Spiegelhalter, 2000](#_ENREF_38)) and linked to the freely available software R (version 4.0.2, Copyright © 2020 ([R foundation for statistical computing, 2014](#_ENREF_47))) through the package R2WinBUGS ([Sturtz, Ligges, & Gelman, 2005](#_ENREF_59)). In all models the MCMC Gibbs sampler was initially run for 10,000 iterations and these were discarded as ‘burn-in’. Models were run for at least further 5,000 iterations, on which inferences were based. Chain convergence was checked using autocorrelation and Brooks-Gelman-Rubin diagram diagnostics ([S. Brooks & Gelman, 1998a](#_ENREF_7); [S. Brooks & Gelman, 1998b](#_ENREF_8); [Gelman & Rubin, 1992](#_ENREF_25)). Goodness of fit and model complexity was assessed using the deviance information criterion (DIC) and posterior mean residual deviance ([Spiegelhalter, Best, Carlin, & van der Linde, 2002](#_ENREF_56)).

We presented the estimated results as relative treatment effect scores and associated 95% credible intervals, CrIs. We have estimated the probability of a treatment being the ‘best’ (i.e. being the most clinically effective) (Ades, 2006), and presented rankograms for all interventions, which provide the probabilities of an intervention being ranked 1 (the most effective) to 7 (the least effective). Finally, we reported the surface under the cumulative ranking curve (SUCRA), which is a numerical presentation of the overall ranking of each intervention. SUCRA values range from 0% to 100%, with higher SUCRA values suggesting that a treatment is likely to be better overall (Dias, 2014; Dias, 2013).

Appendix C gives further details on the analysis (C1), including annotated synthesis WinBUGS code (C2), sample data and initial values for the main model used (C3).

**2.7. Assessment of heterogeneity and consistency**

The model was extended to include study-level covariates as potential treatment effect modifiers. This meant that we looked for factors other than the treatment itself, which could have influenced outcomes within each study and may have created differences (heterogeneity) across studies. These factors included disease severity ([Spitzer, Kroenke, Williams, & Lowe, 2006](#_ENREF_57)), concomitant medication (Berger, Edelsberg, et al., 2011) and the presence of comorbidities (Smith et al., 2010; Pacek et al. 2013). Meta-regression is the most commonly employed method to explore the influence of particular study-level covariates on the relative effect. To preserve all studies (and treatments), when a covariate was not reported by some studies, we allowed the model to impute missing covariate information (multiple imputation procedure assuming ‘missing at random’).

We assessed inconsistency to check that all pieces of evidence (from direct and indirect sources) were in agreement. Following guidance by Dias and colleagues ([Dias, Welton, et al., 2013](#_ENREF_21); [Dias, Welton, Sutton, Caldwell, et al., 2014](#_ENREF_22)), inconsistency was assessed by comparing the DIC of our primary analyses (based on NMA models that assume consistency between direct and indirect evidence) and the DICs yielded by inconsistency models (which provide effect estimates based on direct evidence only). Results were assessed for coherence by qualitatively comparing estimates of pairwise ANCOVA meta-analysis (direct) and ANCOVA RE NMA (direct and indirect).

**2.8. Sensitivity analysis**

We conducted two types of sensitivity analysis. First, we evaluated the sensitivity of the networks to the influence of each individual trial. When network links were informed by more than one trial, we removed each trial one at a time (giving *nj−1* for each analysis, where *n* is the total number of trials in contrast *j*) and investigated the impact on the probability of each intervention being ‘best’. Second, we assessed the robustness of the synthesis results by repeating the analysis while excluding all studies of less than 30 patients.

# RESULTS

**3.1. Included and excluded studies**

Initial systematic searches of bibliographic databases identified 16,272 records; in addition, 32 records were identified through secondary searches (e.g. citation searching of protocols and abstracts). After duplicates were removed, a total of 8,920 records were screened by title and abstract and 8,560 records were excluded. We retrieved the full text papers for the remaining 377 records and, as a result of further screening, 352 articles were excluded. In total, 21 studies (reported in 25 papers) were included in the review. The PRISMA diagram (Fig. 1) summarises the number of records retrieved and selected at different stages of identification and screening. Appendix D gives a full reference list of the excluded studies grouped according to reasons for exclusion.

**Fig. 1 - PRISMA diagram for the identification and selection of clinical trials relating to digital interventions for generalised anxiety disorder (GAD)**

**<Insert Fig 1 here>**

**3.2. Sample characteristics in RCTs of DIs for GAD**

The 21 RCTs included in the review, as detailed in Table 1, were conducted over 10 years between 2009 and 2019 in 10 countries (Sweden, Australia, USA, UK, Canada, Spain, Italy, Ireland, Taiwan, Netherlands) and involved 2,547 randomised participants. Most participants were recruited from the adult general population, except in 4 studies that recruited students/young adults and 1 study with over 60s. GAD populations were defined as either meeting the criteria of an established diagnostic tool, such as the Mini-International Neuropsychiatric Interview (MINI) ([Sheehan et al., 1998](#_ENREF_54)), or a score above an accepted cut-off for diagnosable GAD in standardised questionnaires, such as the Generalised Anxiety Disorder-7 item questionnaire (GAD-7) ([Spitzer, Kroenke, Williams, & Lowe, 2006](#_ENREF_57)).

**Table 1 – Sample characteristics and outcome measurement in RCTs of digital interventions for GAD**

**<Insert Table 1 here>**

**3.3. Risk of bias assessment**

All but one ([Hazen et al., 2009](#_ENREF_27)) out of the 21 included studies, were judged to have a high risk of bias in at least one domain of assessment, for at least one outcome measure. This was largely due to the choice of outcome measurement as all studies used self-reported - albeit standardised – questionnaires. Self-reported outcomes are considered to have a high risk of bias in these studies because participants can rarely be blind to their allocation group. A visual description of the results of the RoB assessment is given in Appendix E, both for each RoB domain across all studies (Appendix E1) and for each study under each RoB domain (Appendix E2).

**3.4. Classification of Digital Interventions and Comparators**

A classification exercise took place to enable consistency across digital interventions and comparators. We classified DIs and their alternatives according to three criteria: a) whether they were a psychological/behavioural intervention (I) or a non-therapeutic psychological/behavioural control (C); b) whether they were digital (D) or non-digital (NoD); c) whether they were supported (S) or unsupported (U). Waiting lists and usual care were classified under no intervention (NI) unless an active component (e.g. monitoring, sham activity) was introduced, in which case the waiting list/usual care was classified as non-therapeutic psychological/behavioural control. An additional classification group was included for pharmacological interventions, called medication (M).

The interventions and controls of the 20 RCTs were allocated to one of the following 8 classification groups: medication (M); no intervention (NI); supported digital control (SDC); supported digital intervention (SDI); supported non-digital control (SNoDC); supported non-digital intervention (SNoDI); unsupported digital control (UDC); unsupported digital intervention (UDI). There were no available clinical studies that included unsupported non-digital interventions (UNoDI) or unsupported non-digital controls (UNoDC). Table 2 describes all the interventions and controls included in each classification group for each study.

**Table 2 – Characteristics and classifications of interventions and controls in RCTs of digital interventions for GAD**

**<Insert Table 2 here>**

Based on the 8-group classification for GAD RCTs, the majority of DIs studied were supported (SDI - 18 RCTs) and were compared against no intervention (NI - 12 RCTs). Only 3 RCTs evaluated unsupported DIs (UDI); two were web-based CBT ([Christensen, Batterham, et al., 2014](#_ENREF_10); [Dear et al., 2015](#_ENREF_16)) and one ([Pham et al.](#_ENREF_46)) a mobile game to practise breathing re-training. The only non-digital intervention (NoDI) represented in two RCTs ([Navarro-Haro et al., 2019](#_ENREF_44); [Topper et al., 2017](#_ENREF_63)) was group therapy (one CBT and one mindfulness-based intervention) and there was only one RCT that included an antidepressant medication, Sertraline ([Christensen, Mackinnon, et al., 2014](#_ENREF_11)). With regards to non-therapeutic active controls reported in 8 RCTs, most included a digital element whereas only one RCT ([Teng et al., 2019](#_ENREF_60)) had a non-digital control in the form of a weekly face-to-face assessment with a research assistant in a lab (SDC).

Just over half of the included RCTs (12/21) evaluated CBT. Therapeutic approaches other than CBT included: psychodynamic therapy (Andersson 2012), extinction therapy (Andersson 2016), acceptance and commitment therapy (Dahlin 2016), cognitive or attentional bias modification (Hazen 2009, Hirsch 2018, Teng 2019), mindfulness (Dahlin 2016, Navarro-Haro 2019), relaxation (Repetto 2013) and diaphragmatic breathing (Pham 2016). The most common technology platform use was a web-interface (n=17). Two studies used VR platforms (Navarro-Haro 2019, Repetto 2013), and two used smartphone apps (Pham 2016, Teng 2019). The study by Repetto et al (2013) also used a mobile interface to enable users to access the VR scenarios at home but has also included a biofeedback system in one of the study arms.

DIs differed not only in technology platform and therapy type, but also in whether additional interpersonal support was offered as an adjunct to the digital element and, if so, its type. Most DIs included some interpersonal contact by phone or face-to-face with professionals (GPs, therapists, psychologists both students and qualified), non-clinical researchers, or lay people. Only two studies included DIs that were pure self-help without any contact (Christensen et al 2014a, Pham 2016). Some DIs were supplemented by standardised emails without regular communication with another person (Christensen 2014a, Dear 2015, Howell 2018).

**3.5. Selection of studies and outcome for the NMA**

A total of 45 different outcome measures were reported in the included RCTs, as shown in Appendix F. GAD-7 was used in 14 out of the 21 RCTs to measure symptoms at baseline and outcomes at follow-up. Penn State Worry Questionnaire (PSWQ) (Meyer, Miller, Metzger, & Borkovec, 1990) was also reported in 14 studies. Apart from GAD-7 and PSWQ, the two other most frequently reported outcomes were for depression: the Patient Health Questionnaire – 9 item (PHQ-9) ([Kroenke & Spitzer, 2002](#_ENREF_35)) and the Beck Depression Inventory – version II (BDI-II) ([Beck, Steer, & Brown, 1996](#_ENREF_6)), reported in 8 and 6 RCTs respectively (Appendix F1). The Hamilton Anxiety Scale (HAM-A) ([Maier, Buller, Philipp, & Heuser, 1988](#_ENREF_39)) used in a recent NMA on medication for GAD ([Slee et al., 2019](#_ENREF_55)), only appeared once in the included RCTs (Appendix F2).

We focused on GAD-7 as our outcome of choice for the NMA. GAD-7 is a 7-item anxiety scale described in the literature as a valid and efficient tool to screen for GAD and assess symptom severity in clinical practice and research ([Spitzer et al., 2006](#_ENREF_57)).

Our NMA for GAD-7 included 13 studies (Table 1). One study ([Howell, Rheingold, Uhde, & Guille, 2018](#_ENREF_31)) used GAD-7 but it was not included in the meta-analysis because it only reported categorical outcomes (i.e. mild, moderate, severe) rather than continuous scores. The measurement period across studies ranged from 3 to 12 weeks, with longer follow-ups only available for very few studies.

Given the high level of reporting of PSWQ, data for this outcome was also quantitatively synthesised (see Appendix J).

**3.6. NMA results for GAD-7 scores at follow-up**

Ten direct treatment comparisons were made in the 13 trials included in the GAD-7-based NMA; 4 of the 13 trials were multi-arm trials (three 3-arm trials ([Christensen, Mackinnon, et al., 2014](#_ENREF_11); [Repetto et al., 2013](#_ENREF_48); [Robinson et al., 2010](#_ENREF_53)) and one 5-arm trial ([Christensen, Batterham, et al., 2014](#_ENREF_10))); 5 comparisons were informed by more than one trial where pair-wise ANCOVA meta-analysis was conducted (ANCOVA FE models and ANCOVA RE for when n>3).

We constructed a network plot to illustrate which interventions had been compared head-to-head (direct pairwise comparisons) for GAD-7 within the 13 included RCTs. An overview of these pairwise comparisons and synthesised data are shown in Appendix G. The structure of the network for GAD-7 is shown in Figure 2.

**Fig. 2 - Network plot for comparisons between all interventions and controls for GAD populations in RCTs with GAD-7 scores**

**< Insert Figure 2 here >**

Fixed- and random-effects models were employed with minimal difference in mean residual deviances and DIC identified between the models tested. However, posterior estimates of between-study heterogeneity, $τ^{2}$, suggested considerable variability across studies, which was in line with the narrative assessment of the studies. Hence, a random-effects approach was preferred. There was a high degree of uncertainty in the network results, especially in links not informed by direct evidence. Table 3 presents the full results of the NMA based on GAD-7 scores.

**Table 3 - Full meta-analysis results: network and direct pairwise comparisons between all interventions and controls for post-treatment (12 weeks) GAD-7 scores adjusted for baseline**

**<Insert Table 3 here>**

Medication (M) was associated with the largest decrease in GAD-7 median scores compared to the other interventions, although uncertainty was high in the NMA estimates, with all 95% credible intervals including zero. These results are driven by the outcomes of a small (n=21), three-arm, trial ([Christensen, Mackinnon, et al., 2014](#_ENREF_11)) that compared medication supplemented with scheduled face-to-face meetings with psychologists and GPs, against SDC (a general health website with scheduled meetings with psychologists & GPs) and SDI (a web-based CBT self-help programme with scheduled meetings with psychologists & GPs). The adjustment for baseline scores indicated that the baseline effect on the final outcome is small with a 95% credible interval including zero (change in GAD score: -0.14 [95% CrI -1.10 to 0.82]).

Results of independently pooling direct evidence for each contrast (but not pooling when n=1) were found to be generally consistent with the NMA results, both in terms of direction and magnitude of the estimates (Table 3, upper-right triangle, shaded). Of note are the differences in the estimates found when applying fixed- and random-effects ANCOVA meta-analysis model on direct evidence for the comparisons of SDIs *vs* SDCs (n=4) and SDIs *vs* NI (n=8), evidencing non-negligible variability across studies and the importance of accounting for between-study heterogeneity.

Based on SUCRA values and rankograms for each intervention, as detailed in Appendices H1 and H2 respectively, SDIs were estimated to be more effective (i.e. ranked higher) than UDIs, which included unsupported web-based CBT ([Christensen, Batterham, et al., 2014](#_ENREF_10); [Dear et al., 2015](#_ENREF_16)) and an unsupported mobile breathing retraining game ([Pham et al., 2016](#_ENREF_46)); however SDIs were less effective than SNoDI, a weekly group mindfulness-based intervention with a therapist ([Navarro-Haro et al., 2019](#_ENREF_44)).

Similar analysis was performed on the PSWQ outcome. Results are shown in Appendix J.

**3.8. Results of between-study heterogeneity and inconsistency assessments**

Three sources of heterogeneity were considered relevant: disease severity, concomitant medication and comorbidities. Using data relating to disease severity and comorbidities was not feasible – see Appendix I for further details – thus only data on concomitant medication was included as a covariate in the synthesis modelling. When this covariate is included, the between-study heterogeneity parameter, $τ^{2}$, is not reduced, suggesting that heterogeneity is not explained by this covariate. Crucially, even if the proportion receiving concomitant medication was identified as an important effect modifier, the meta-regression model is not necessarily suited to detect this intervention-covariate interaction as patients were receiving medication before trial entry. Therefore, medication may have already exerted an effect on patients, being captured by the ANCOVA baseline adjustment component.

Several data loops existed in the network, where both direct and indirect data informed intervention effectiveness estimates; the possibility of inconsistencies was investigated. Table 3 showed no evidence of substantial discrepancies between the direct and the NMA results; given the uncertainty in the data, only very large differences were likely to result in statistical significance. Results of the consistency and inconsistency models indicated the existence of overall model consistency, as detailed in Appendix I.

Similar analysis was performed on the PSWQ outcome. Results are shown in Appendix J.

**3.9. Sensitivity analysis results**

The sensitivity of the network to specific studies was investigated. In total, 10 analyses with 12 (rather than the total 13) included studies for GAD-7 were performed, and the probability of each intervention being the best was assessed. The SSRI and group CBT continued to have the highest chances of being ‘best’, with probabilities of around 43% and 30%, respectively.

Two studies ([Christensen, Mackinnon, et al., 2014](#_ENREF_11); [Repetto et al., 2013](#_ENREF_48)) had less than 30 patients. Excluding these studies from the network also removed medication (M) from the comparator set, altering the network structure. As expected, with the reduction in the number of studies informing the network, the uncertainty in the posterior effect distributions increased further. However, no significant changes were observed compared to the main model results.

Similar analysis was performed on the PSWQ outcome. The ranking of active interventions in terms of median PSWQ score decrease vs no intervention (NI) was unaltered, although higher score decreases were estimated. Comparing the direction and magnitude of differences in median scores at follow-up between GAD-7 and PSWQ results (where available for both), we make three observations (Appendix K). First, the difference in GAD-7 median scores at follow-up between medication and DIs is the largest across all comparators and favours medication. Second, there were no data available for comparisons between DIs and individual therapy, either face-to-face or by telephone, or between DIs and manualised guided self-help (which is the non-digital counterpart of most DIs). Third, the direction of effect favoured SDIs for GAD-7 and UDIs for PSWQ.

# DISCUSSION

**4.1. Summary and interpretation**

Our systematic review retrieved 21 RCTs of DIs or alternative pathways of care, including no intervention, for GAD. Comparators included in the studies varied. Specifically, interventions or controls could be digital or non-digital and supported or unsupported by clinicians or lay people. The majority of comparisons were between supported digital interventions and no intervention. Using an ANCOVA framework, our main NMA on GAD-7 pooled together post-treatment scores - adjusted for baseline. In addition, the existence of treatment effect modifiers was assessed, several sensitivity analyses were carried out and network consistency evaluated. NMA on the PSWQ outcome was also performed.

Our NMA results suggest that medication is associated with lower anxiety scores at follow-up relative to all other interventions and controls. Medication also ranks first in terms of its likelihood of being most effective, which considers the uncertainty in relative effect estimates. Medication results are based on data from one study. Antidepressant medication as a treatment for GAD is supported by clinical guidelines ([National Institute for Health and Clinical Excellence, 2011](#_ENREF_43)) and previous evidence syntheses. A large NMA ([Slee et al., 2019](#_ENREF_55)) of medication against placebo for GAD found that Sertaline (the same antidepressant used in the study by Christensen and colleagues included in our NMA) improved HAM-A scores by a mean difference of –2·88 (CrI –4·17 to –1·59) from baseline compared to a placebo based on 6 trials. Another meta-analysis involving a mixed population ([Cuijpers et al., 2014](#_ENREF_14)) favoured a combined treatment of psychological therapy and medication for all depressive and anxiety disorders, except GAD, where the direction of effect favoured antidepressant medication (Venlafaxine) alone.

Previous reviews of DIs that reported GAD-related outcomes ([Cuijpers et al., 2014](#_ENREF_14); [Richards, Richardson, Timulak, & McElvaney, 2015](#_ENREF_50)) used mixed samples of anxiety disorders and depression, without reporting outcomes separate for GAD subgroups. There are no RCTs with GAD populations comparing DIs with non-digital self-help interventions based on a manual rather than a web-based program. Also, no RCTs compare DIs with individual therapy for GAD, either face-to-face or by telephone; the only available comparisons in the literature are between DIs and group therapy.

Due to very wide confidence intervals, our NMA results were inconclusive as to whether DIs for GAD were better than no intervention or non-therapeutic active controls, or whether they confer an additional benefit to standard therapy. Previous meta-analyses have suggested that supported DIs could be as good as face-to-face therapy across depression, anxiety and somatic disorders ([Cuijpers, Donker, van Straten, Li, & Andersson, 2010](#_ENREF_13); Carlbring et al 2018). However, the mixed samples in these meta-analyses without separate analysis or reporting for GAD sub-samples does not allow any conclusions about the relative efficacy of DIs specific to the treatment and prevention of GAD.

The results for supported vs unsupported DIs for GAD were counterintuitive, as we would expect supported DIs to rank higher in terms of the likelihood of being ‘best’, based on a previous meta-analysis in which supported DIs were found to be four times more likely to be effective compared to those without any therapist contact (Spek et al., 2007). We found that unsupported DIs rank higher than supported DIs in terms of the probability of being best, but vice versa when considering all rankings (SUCRAs). This is consistent with a recent review (Shim et al 2017) that reported mixed findings regarding guided versus unguided DIs and human versus automated support for DIs. This suggests that the design, content, technology platform or type of reinforcement offered in lieu of personal support in unsupported DIs may be important and account for some of the variability in outcomes.

**4.2. Strengths and limitations**

To our knowledge this is the first ANCOVA NMA model to synthesise evidence on two widely used outcomes for GAD. Our NMA makes best use of all currently available RCT-based evidence on DIs for GAD. Despite the sparse and low-quality data, a statistical synthesis can still be useful for decision-makers in mental health (including healthcare professionals, providers and policy-makers, patients and their families, and the research community) who may be considering the use of DIs for GAD, so that they are properly informed about the current status of the evidence base, know which DIs have been shown to be more effective in reducing GAD and prioritise future research.

There was substantial uncertainty around effect estimates of DIs against alternatives for GAD-7. This is driven by the small number of studies informing most comparisons, the small sample sizes used in some of these studies and their high risk of bias across the evidence base, all limiting our confidence in any observed differences in anxiety scores between intervention, comparators and control arms. These observed differences may simply be due to chance; but in view of the current evidence base we cannot make clear recommendations about the relative effectiveness of DIs against their comparators.

We have to use caution when interpreting the results of our NMA across all different interventions for GAD. Our review has been completed in the context of DIs; it only included RCTs in which at least one of the randomisation arms was a DI. Therefore, we cannot draw any conclusions about the comparative merit of non-digital interventions (psychological or pharmacological) for GAD when these are considered separately to DIs (for example group CBT vs. medication). To be able to do this, we would need to include RCTs in an NMA that would enable 2nd or 3rd order contrasts (e.g. RCTs comparing no intervention and medication), which was beyond the scope of this review. Also, ranking based on likelihood of being best and on SUCRAs does not reflect differences in effectiveness estimates between interventions and controls and credible intervals, that is, we cannot tell whether the differences between ranking position (e.g. between 1st, 2nd, 3rd, 4th, etc) are clinically meaningful.

Another point of caution, as with all evidence synthesis of complex interventions, is the pooling DIs and their alternatives into groups for analysis based on our classification criteria. Any classification implies interpretation and judgement which is conditional upon the information available from included studies. We note the insufficient reporting of details about ‘non-therapeutic controls’ and waiting list in some studies. Furthermore, we could have split DIs into further categories according to the technology used (e.g. VR, internet, mobile app), or the function of the technology (e.g. adjunct to clinician-delivered therapy vs. patient self-help), or the type of support (e.g. phone calls vs meetings). This would have created more “nodes” in the NMA models, but also more uncertainty because comparisons between DIs and their alternatives within each subgroup would have been informed by fewer studies.

Many of the included RCTs recruited small samples and involved multiple arms, often comparing different versions of the same intervention, thereby reducing the power of the study. Our evidence synthesis also shows that the majority of RCTs have either a short timeframe for follow-up (up to 12 weeks), or the control group has already crossed to the intervention at the point of a longer follow-up (up to 2 years), which undermines the original randomisation. Consequently, we did not include observations for further follow-up time points, where these were available, nor did we account for time differences in the short-term outcome reporting (post- treatment assessments varied from 3 to 12 weeks). Our NMA results reflect the short-term impact of DIs over an initial treatment period, but there is scant evidence to inform randomised comparisons about effectiveness beyond 12 weeks.

**4.3. Recommendations**

As GAD is the most prevalent and least studied condition among other common mental health problems, future evidence syntheses will be helpful to focus on GAD populations and stratified GAD subgroups where these are randomised within mixed populations, as means of informing GAD-specific future research and clinical guidelines ([National Institute for Health and Clinical Excellence, 2011](#_ENREF_43)). Feasibility and pilot studies, as well as user involvement in the development of the intervention and delivery protocols, could ensure that the final RCT tests the best possible intervention for GAD. Adaptive designs with improved intervention features and boosted recruitment numbers to a fully powered RCT are preferable to the underpowered studies with multiple arms testing increments of the same DI that we have identified.

Our NMAs and previous literature suggest that antidepressants are an important factor to consider in future studies on DIs for GAD. Psychological interventions – whether digital or non-digital – include participants who are taking medication as part of routine care. It is difficult to disentangle the effects of medication and psychological support for GAD, and future RCTs need to report medication details (name, dose and duration) and include it as a covariate in their analysis to establish how outcomes with DIs and controls are influenced by concurrent medication use.

The evidence base available in this setting is complex. In particular, the sheer volume of anxiety metrics (45 in total) being reported across the available studies, suggests a lack of consensus on which measures to use in evaluating GAD outcomes. Having a consensus about GAD-specific outcome measures can prevent participant fatigue from completing batteries of different questionnaires and enable comparisons across studies and data synthesis. GAD-7 is more sensitive to changes associated with treatment and therefore may be more suitable for longitudinal clinical research (Dear et al., 2011). Reporting continuous data on the GAD-7 as a common measure in RCTs with GAD populations will make more studies available for a future statistical synthesis. Including HAM-A in studies of psychological therapies will enable us to compare results with pharmacological studies. Future analyses using multivariate models may be able to make better use of the available evidence by borrowing strength across different outcomes.

Many studies that follow up participants over the longer-term offer the intervention to those randomised to the control group at a crossover point. Participants are also likely to receive some treatment as part of usual care the longer they remain on waiting lists or non-therapeutic controls, so studies cannot withhold interventions to enable long-term follow-up. As the typical duration of DIs is between 3-to-12 weeks, the follow-up period of future RCTs needs to be longer, for example 6 months, to help us better understand the “stickiness” longer term effects of DIs beyond their initial delivery period. Usual care and waiting lists are poorly reported in RCTs and do not include data on concurrent interventions accessed by participants, including openly available self-help, which can influence the observed difference in outcome between DIs and no intervention. Greater clarity and more detailed reporting about the specific elements of comparators is essential to improve our understanding of the effects of DIs.

# CONCLUSIONS

This study is the first to evaluate the effectiveness of DIs specifically in a GAD population. It is also the first to combine all the RCT-based effectiveness evidence from DIs and key comparators in a single modelling framework, allowing the estimation of relative treatment effects for all relevant comparisons. Our results suggest that antidepressant medication is associated with lower anxiety scores at follow-up relative to all other interventions and controls. Results were inconclusive as to whether DIs are better than no intervention and non-therapeutic active controls for GAD, or whether they confer an additional benefit to standard therapy. Overall, our findings are limited in informing decision-making, highlighting how little is currently known about the comparative effectiveness of such interventions. Future primary studies and meta-analyses need to focus on GAD populations rather than mixed samples, or report outcomes specifically for GAD sub-samples if they intend to answer questions about the comparative effectiveness of DIs for GAD. Comparing DIs with manualised (non-digital) self-help and individual therapy, for which there are no current RCTs for GAD populations, will be useful in the context of stepped care. Antidepressant medication for GAD as a first-line treatment against DIs deserves further research and economic modelling. To inform commissioning and potential disinvestment from non-digital alternatives, we need to put the findings of this evidence synthesis into context together with an assessment of the costs of developing and implementing DIs in clinical practice.

# ACKNOWLEDGEMENTS

We thank Prof Paul McNamee, Prof Nicola Cooper, Dr Alrinda Cerga-Pashoja and Mr Daniel Horne for their guidance and support as members of our advisory/steering group.

**Disclaimer:** This paper presents independent research supported by the National Institute for Health Research (NIHR), Health Technology Assessment (HTA) Programme. The views expressed are those of the authors and not necessarily of the National Health Service, the NIHR, or the Department of Health. The content of this manuscript has previously appeared online as a preprint of the NIHR report: https://eprints.whiterose.ac.uk/171555/.

# AUTHORS CONTRIBUTIONS

Pedro Saramago reviewed the clinical evidence, performed the data synthesis and statistical analysis, interpreted findings, wrote the manuscript and led the work as a whole; Lina Gega reviewed the clinical evidence, classified the interventions, interpreted findings, supported the write up of the manuscript and all aspects of the review; David Marshall identified and selected the studies, extracted study data, assessed their risk of bias, interpreted findings and reviewed the manuscript; Georgios F Nikolaidis supported the data synthesis and statistical analysis and reviewed the manuscript; Dina Jankovic contributed to the selection of studies, interpreted findings and reviewed the manuscript; Hollie Melton identified and selected the studies, extracted study data, assessed their risk of bias, interpreted findings and reviewed the manuscript; Sarah Dawson performed the searches and reviewed the manuscript; Rachell Churchill provided support on all aspects of the review and reviewed the manuscript; and Laura Bojke provided support on all aspects of the review and reviewed the manuscript.

# FUNDING

This work was supported by the UK’s National Institute for Health Research (NIHR), Health Technology Assessment (HTA) Programme, grant number 17/93/06, 2019.

# DECLARATION OF INTERESTS

Lina Gega is the fund holder of the research grant that supported this work. Pedro Saramago, David Marshall, Georgios Nikolaidis, Dina Jankovic, Hollie Melton, Sarah Dawson, Rachel Churchill and Laura Bojke declare that they have no conflict of interest.

# REFERENCES

# Aboujaoude, E., Gega, L., Parish, M. B., & Hilty, D. M. (2020). Editorial: Digital interventions in mental health: Current status and future directions. [Editorial]. Frontiers in Psychiatry, 11(111). doi: 10.3389/fpsyt.2020.00111

# Ades, A. E., Sculpher, M., Sutton, A., Abrams, K., Cooper, N., Welton, N., & Lu, G. (2006). Bayesian methods for evidence synthesis in cost-effectiveness analysis. Pharmacoeconomics, 24(1), 1-19. doi: 10.2165/00019053-200624010-00001

# Andersson, E., Hedman, E., Wadstrom, O., Boberg, J., Andersson, E. Y., Axelsson, E., . . . Ljotsson, B. (2016). Internet-Based Extinction Therapy for Worry: A Randomized Controlled Trial. [In Press]. Behavior Therapy., 02. doi: http://dx.doi.org/10.1016/j.beth.2016.07.003

# Andersson, G., Paxling, B., Roch-Norlund, P., Ostman, G., Norgren, A., Almlov, J., Silverberg, F. (2012). Internet-based psychodynamic versus cognitive behavioral guided self-help for generalized anxiety disorder: A randomized controlled trial. Psychother Psychosom, 81(6), 344-355. doi: 10.1159/000339371

# Barlow, D. H. (2002). Anxiety and its Disorders : The Nature and Treatment of Anxiety and Panic (2nd ed.). New York: Guilford Press.

# Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck Depression Inventory-II. San Antonio, 78(2), 490-498.

# Berger, A., Edelsberg, J., Bollu, V., Alvir, J. M. J., Dugar, A., Joshi, A. V., et al. Healthcare utilization and costs in patients beginning pharmacotherapy for generalized anxiety disorder: a retrospective cohort study. BMC Psych. 2011; 11(193). PubMed Central PMCID: PMC22151689

# Brooks, S., & Gelman, A. (1998a). General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics, 7(4), 434-455. doi: Doi 10.2307/1390675

# Brooks, S., & Gelman, A. (1998b). Some issues in monitoring convergence of iterative simulations. Dimension Reduction, Computational Complexity and Information, 30, 30-36.

# Caldwell, D. M., Ades, A. E., & Higgins, J. P. (2005). Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ, 331(7521), 897-900. doi: 10.1136/bmj.331.7521.897

# Carlbring, P., Andersson, G., Cuijpers, P., Riper, H., & Hedman-Lagerlöf, E. (2018). Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: An updated systematic review and meta-analysis. Cognitive Behaviour Therapy (Vol. 47, Issue 1, pp. 1–18).

# Christensen, H., Batterham, P., Mackinnon, A., Griffiths, K. M., Hehir, K. K., Kenardy, J., . . . Bennett, K. (2014). Prevention of generalized anxiety disorder using a web intervention, iChill: Randomized controlled trial. Journal of Medical Internet Research, 16(9). doi: 10.2196/jmir.3507

# Christensen, H., Mackinnon, A. J., Batterham, P. J., O'Dea, B., Guastella, A. J., Griffiths, K. M., . . . Hickie, I. (2014). The effectiveness of an online e-health application compared to attention placebo or Sertraline in the treatment of Generalised Anxiety Disorder. Internet Interventions, 1(4), 169-174. doi: https://doi.org/10.1016/j.invent.2014.08.002

# Cooper, N. J., Peters, J., Lai, M. C., Juni, P., Wandel, S., Palmer, S., . . . Sutton, A. J. (2011). How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? Value Health, 14(2), 371-380. doi: 10.1016/j.jval.2010.09.001

# Cuijpers, P., Donker, T., van Straten, A., Li, J., & Andersson, G. (2010). Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med, 40(12), 1943-1957. doi: 10.1017/S0033291710000772

# Cuijpers, P., Sijbrandij, M., Koole, S. L., Andersson, G., Beekman, A. T., & Reynolds III, C. F. (2014). Adding psychotherapy to antidepressant medication in depression and anxiety disorders: A meta-analysis. World Psychiatry, 13(1), 56-67. doi: 10.1002/wps.20089

# Dahlin, M., Andersson, G., Magnusson, K., Johansson, T., Sjogren, J., Hakansson, A., . . . Carlbring, P. (2016). Internet-delivered acceptance-based behaviour therapy for generalized anxiety disorder: A randomized controlled trial. [Cognitive Therapy 3311]. Behaviour Research and Therapy, 77, 86-95. doi: http://dx.doi.org/10.1016/j.brat.2015.12.007

# Dear, B. F., Staples, L. G., Terides, M. D., Karin, E., Zou, J., Johnston, L., . . . Titov, N. (2015). Transdiagnostic versus disorder-specific and clinician-guided versus self-guided internet-delivered treatment for generalized anxiety disorder and comorbid disorders: A randomized controlled trial. [Cognitive Therapy 3311]. Journal of Anxiety Disorders, 36, 63-77. doi: http://dx.doi.org/10.1016/j.janxdis.2015.09.003

# Dear, B. F., Titov, N., Sunderland, M., McMillan, D., Anderson, T., Lorian, C., & Robinson, E. (2011). Psychometric comparison of the Generalized Anxiety Disorder Scale-7 and the Penn State Worry Questionnaire for measuring response during treatment of generalised anxiety disorder. Cogn Behav Ther, 40(3), 216-227. doi: 10.1080/16506073.2011.582138

# Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2008). Chapter 9: Analysing data and undertaking meta-analyses. In Higgins J. P. T. & G. S. (Eds.), Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.1 [updated September 2008]: The Cochrane Collaboration.

# Dias, S., Sutton, A. J., Ades, A. E., & Welton, N. J. (2013). Evidence synthesis for decision making 2: A generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Medical Decision Making, 33(5), 607-617. doi: 10.1177/0272989x12458724

# Dias, S., Welton, N. J., Sutton, A. J., & Ades, A. E. (2014). NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. London.

# Dias, S., Welton, N. J., Sutton, A. J., Caldwell, D. M., Lu, G., & Ades, A. E. (2013). Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. Med Decis Making, 33(5), 641-656. doi: 10.1177/0272989X12455847

# Dias, S., Welton, N. J., Sutton, A. J., Caldwell, D. M., Lu, G., & Ades, A. E. (2014). NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials. London.

# Fu, R., Vandermeer, B. W., Shamliyan, T. A., O'Neil, M. E., Yazdi, F., Fox, S. H., & Morton, S. C. (2008). Handling continuous outcomes in quantitative synthesis Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville (MD).

# Gega, L., & Gilbody, S. (2015). Software-based psychotherapy: The example of computerized cognitive-behavioral therapy Mental Health in the Digital Age: Grave dangers, great promise. (pp. 196-219). Oxford: Oxford University Press.

# Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. Statistical Science, 7(4), 457-472. doi: 10.2307/2246093

# Hayes-Skelton, S. A., Roemer, L., Orsillo, S. M., & Borkovec, T. D. (2013). A contemporary view of applied relaxation for generalized anxiety disorder. Cognitive behaviour therapy, 42(4), 292-302.

# Hazen, R. A., Vasey, M. W., & Schmidt, N. B. (2009). Attentional retraining: a randomized clinical trial for pathological worry. Journal of Psychiatric Research, 43(6), 627-633.

# Higgins, J. P. T., & Cochrane Collaboration. (2020). Cochrane handbook for systematic reviews of interventions (Second edition. ed.). Hoboken, NJ: Wiley-Blackwell.

# Hirsch, C. R., Krahe, C., Whyte, J., Loizou, S., Bridge, L., Norton, S., & Mathews, A. (2018). Interpretation training to target repetitive negative thinking in generalized anxiety disorder and depression. Journal of Consulting and Clinical Psychology, 86(12), 1017-1030. doi: http://dx.doi.org/10.1037/ccp0000310

# Hodgins, D. C., & Peden, N. (2008). Cognitive-behavioral treatment for impulse control disorders. Revista Brasileira de Psiquiatria, 30(Suppl1), S31-S40. doi: 10.1590/S1516-44462006005000055

# Howell, A. N., Rheingold, A. A., Uhde, T. W., & Guille, C. (2018). Web-based CBT for the prevention of anxiety symptoms among medical and health science graduate students. Cognitive Behaviour Therapy, 1-21. doi: https://dx.doi.org/10.1080/16506073.2018.1533575

# Hunot, V., Churchill, R., Teixeira, V., & de Lima, M. S. (2007). Psychological therapies for generalised anxiety disorder. Cochrane Database of Systematic Reviews(1).

# Johansson, R., Bjorklund, M., Hornborg, C., Karlsson, S., Hesser, H., Ljotsson, B., . . . Andersson, G. (2013). Affect-focused psychodynamic psychotherapy for depression and anxiety through the Internet: A randomized controlled trial. PeerJ, 1, e102. doi: https://dx.doi.org/10.7717/peerj.102

# Jones, S. L., Hadjistavropoulos, H. D., & Soucy, J. N. (2016). A randomized controlled trial of guided internet-delivered cognitive behaviour therapy for older adults with generalized anxiety. [Cognitive Therapy 3311]. Journal of Anxiety Disorders, 37, 1-9. doi: http://dx.doi.org/10.1016/j.janxdis.2015.10.006

# Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: a new depression diagnostic and severity measure. Psychiatric annals, 32(9), 509-515.

# Latimer NR, Abrams KR, Lambert PC, Crowther MJ, Wailoo AJ, Morden JP, Akehurst RL, Campbell MJ. (2014) Adjusting survival time estimates to account for treatment switching in randomized controlled trials--an economic evaluation context: methods, limitations, and recommendations. Med Decis Making. 34(3):387-402.

# Lu, G., & Ades, A. E. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med, 23(20), 3105-3124. doi: 10.1002/sim.1875

# Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. Stat Med, 21(16), 2313-2324. doi: 10.1002/sim.1201

# Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Statistics and Computing, 10(4), 325-337. doi: Doi 10.1023/A:1008929526011

# Maier, W., Buller, R., Philipp, M., & Heuser, I. (1988). The Hamilton Anxiety Scale: reliability, validity and sensitivity to change in anxiety and depressive disorders. Journal of affective disorders, 14(1), 61-68.

# Mental Health Foundation. (2016). Fundamental Facts About Mental Health

# 2016. Mental Health Foundation: London.

# McManus, S., Bebbington, P., Jenkins, R., & Brugha, T. (2016). Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014. A survey carried out for NHS Digital by NatCen Social Research and the Department of Health Sciences. University of Leicester.

# Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. Behav Res Ther, 28(6), 487-495. doi: 10.1016/0005-7967(90)90135-6

# Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine, 151(4), 264-269.

# National Institute for Health and Clinical Excellence. (2011). Generalised Anxiety Disorder and Panic Disorder in Adults: Management. Clinical Guideline 113. London: NICE.

# Navarro-Haro, M. V., Modrego-Alarcon, M., Hoffman, H. G., Lopez-Montoyo, A., Navarro-Gil, M., Montero-Marin, J., . . . Garcia-Campayo, J. (2019). Evaluation of a mindfulness-based intervention with and without virtual reality dialectical behavior therapy mindfulness skills training for the treatment of generalized anxiety disorder in primary care: A pilot study. [Behavior Therapy & Behavior Modification 3312]. Frontiers in Psychology, 10. doi: http://dx.doi.org/10.3389/fpsyg.2019.00055

# Pacek, L. R., Storr, C. L., Mojtabai, R., Green, K. M., La Flair, L. N., Alvanzo, A. A., Cullen, B. A., & Crum, R. M. (2013). Comorbid Alcohol Dependence and Anxiety Disorders: A National Survey. Journal of dual diagnosis, 9(4), 10.1080/15504263.2013.835164. https://doi.org/10.1080/15504263.2013.835164

# Paxling, B., Almlov, J., Dahlin, M., Carlbring, P., Breitholtz, E., Eriksson, T., & Andersson, G. (2011). Guided internet-delivered cognitive behavior therapy for generalized anxiety disorder: A randomized controlled trial. [Cognitive Therapy 3311]. Cognitive Behaviour Therapy, 40(3), 159-173. doi: http://dx.doi.org/10.1080/16506073.2011.576699

# Pham, Q., Khatib, Y., Stansfeld, S., Fox, S., & Green, T. (2016). Feasibility and efficacy of an mHealth game for managing anxiety: "Flowy" randomized controlled pilot trial and design evaluation. Games for Health Journal, 5(1), 50-67. doi: https://dx.doi.org/10.1089/g4h.2015.0033

# R foundation for statistical computing. (2014). R version 3.6.0.

# Repetto, C., Gaggioli, A., Pallavicini, F., Cipresso, P., Raspelli, S., & Riva, G. (2013). Virtual reality and mobile phones in the treatment of generalized anxiety disorders: A phase-2 clinical trial. Personal and Ubiquitous Computing, 17(2), 253-260. doi: 10.1007/s00779-011-0467-0

# Revicki, D. A., Travers, K., Wyrwich, K. W., Svedsäter, H., Locklear, J., Mattera, M. S., . . . Montgomery, S. (2012). Humanistic and economic burden of generalized anxiety disorder in North America and Europe. Journal of affective disorders, 140(2), 103-112.

# Richards, D., Richardson, T., Timulak, L., & McElvaney, J. (2015). The efficacy of internet-delivered treatment for generalized anxiety disorder: A systematic review and meta-analysis. Internet Interventions, 2(3), 272-282.

# Richards, D., Timulak, L., Rashleigh, C., McLoughlin, O., Colla, A., Joyce, C., . . . Anderson-Gibbons, M. (2016). Effectiveness of an internet-delivered intervention for generalized anxiety disorder in routine care: A randomised controlled trial in a student population. Internet Interventions, 6, 80-88. doi: https://dx.doi.org/10.1016/j.invent.2016.10.003

# Riley, R. D., Kauser, I., Bland, M., Thijs, L., Staessen, J. A., Wang, J., . . . Deeks, J. J. (2013). Meta-analysis of randomised trials with a continuous outcome according to baseline imbalance and availability of individual participant data. Stat Med, 32(16), 2747-2766. doi: 10.1002/sim.5726

# Robinson, E., Titov, N., Andrews, G., McIntyre, K., Schwencke, G., & Solley, K. (2010). Internet treatment for generalized anxiety disorder: A randomized controlled trial comparing clinician vs. technician assistance. Plos One, 5(6). doi: 10.1371/journal.pone.0010942

# Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., . . . Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (MINI): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. The Journal of Clinical Psychiatry(59(Suppl 2)), 22–33.

# Shim M, Mahaffey B, Bleidistel M, Gonzalez A. (2017 A scoping review of human-support factors in the context of Internet-based psychological interventions (IPIs) for depression and anxiety disorders.Clinical Psychology Review, 57:129-140.

# Slee, A., Nazareth, I., Bondaronek, P., Liu, Y., Cheng, Z., & Freemantle, N. (2019). Pharmacological treatments for generalised anxiety disorder: A systematic review and network meta-analysis. The Lancet, 393(10173), 768-777.

# Smith, J. P., & Book, S. W. (2010). Comorbidity of generalized anxiety disorder and alcohol use disorders among individuals seeking outpatient substance abuse treatment. Addictive behaviors, 35(1), 42–45. https://doi.org/10.1016/j.addbeh.2009.07.002

# Spek, V., Cuijpers, P., Nyklícek, 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.

# Spiegelhalter, D. J., Best, N. G., Carlin, B. R., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B-Statistical Methodology, 64, 583-616. doi: Doi 10.1111/1467-9868.00353

# Spitzer, R. L., Kroenke, K., Williams, J. B., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. Arch Intern Med, 166(10), 1092-1097. doi: 10.1001/archinte.166.10.1092

# Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., . . . Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. BMJ, 366, l4898. doi: 10.1136/bmj.l4898

# Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. Journal of Statistical Software, 12(3), 1-16.

# Teng, M.-H., Hou, Y.-M., Chang, S.-H., & Cheng, H.-J. (2019). Home-delivered attention bias modification training via smartphone to improve attention control in sub-clinical generalized anxiety disorder: A randomized, controlled multi-session experiment. Journal of Affective Disorders, 246, 444-451. doi: 10.1016/j.jad.2018.12.118

# Titov, N., Andrews, G., Johnston, L., Robinson, E., & Spence, J. (2010). Transdiagnostic Internet treatment for anxiety disorders: A randomized controlled trial. [Cognitive Therapy 3311]. Behaviour Research and Therapy, 48(9), 890-899. doi: http://dx.doi.org/10.1016/j.brat.2010.05.014

# Titov, N., Andrews, G., Robinson, E., Schwencke, G., Johnston, L., Solley, K., & Choi, I. (2009). Clinician-assisted Internet-based treatment is effective for generalized anxiety disorder: Randomized controlled trial. [Cognitive Therapy 3311]. Australian and New Zealand Journal of Psychiatry, 43(10), 905-912. doi: http://dx.doi.org/10.1080/00048670903179269

# Topper, M., Emmelkamp, P. M., Watkins, E., & Ehring, T. (2017). Prevention of anxiety disorders and depression by targeting excessive worry and rumination in adolescents and young adults: A randomized controlled trial. Behav Res Ther, 90, 123-136. doi: 10.1016/j.brat.2016.12.015

# van Breukelen, G. J. (2013). ANCOVA versus CHANGE from baseline in nonrandomized studies: The difference. Multivariate Behav Res, 48(6), 895-922. doi: 10.1080/00273171.2013.831743

# Vickers, A. J. (2001). The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: A simulation study. BMC Med Res Methodol, 1, 6. doi: 10.1186/1471-2288-1-6

# Vickers, A. J., & Altman, D. G. (2001). Statistics notes: Analysing controlled trials with baseline and follow up measurements. BMJ, 323(7321), 1123-1124. doi: 10.1136/bmj.323.7321.1123

# Winkens, B., van Breukelen, G. J., Schouten, H. J., & Berger, M. P. (2007). Randomized clinical trials with a pre- and a post-treatment measurement: Repeated measures versus ANCOVA models. Contemp Clin Trials, 28(6), 713-719. doi: 10.1016/j.cct.2007.04.002

# World Health Organization. (2018). International classification of diseases, 11th revision (ICD-11). Geneva: World Health Organization.

World Health Organization. (2019). WHO guideline recommendations on digital interventions for health system strengthening. Geneva: World Health Organization.

# Tables

**Table 1 – Sample characteristics and outcome measurement in RCTs of digital interventions for GAD**

| **Study** | **Country** | **Population** | **Entry criteria**  | **N randomised** | **N analysed** | **Follow-up time points** | **Reported outcomes++** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Andersson 2012 | Sweden | General - adults | GAD diagnosis on SCID-I | 81 | 59 | 8 weeks3 months18 months\* | PSWQ, BAI. BDI-II, GADQ-IV, MADRS-S, QOLI, SCID-I, STAI-S, STAI-T |
| Andersson 2016 | Sweden | General - adults | PSWQ>56 | 140 | 132 | 10 weeks4 months\*12 months\* | PSWQ, BBQ, CAQ, HADS-A, IOU, MADRS-S, MCQ-30 |
| Christensen 2014a | Australia | General - adults | GAD7>5 | 558 | 264 | 10 weeks6 months12 months | GAD-7, PSWQ, ASI, CES-D, days out of role, MINI |
| Christensen 2014b | Australia | General - adults | GAD diagnosis on ADIS-IV | 21 | 11 | 10 weeks6 months12 months | GAD-7, CES-D, CGI |
| Dahlin 2016 | Sweden | General - adults | GAD diagnosis on SCID-IPSWQ>45 | 103 | 85 | 9 weeks6 months\* | GAD-7, PSWQ, BAI, GADQ-IV, MADRS-S, PHQ-9, QOLI |
| Dear 2015 | Australia | General - adults | GAD diagnosis on MINIGAD7>5 | 338 | 260 | 9 weeks3 months12 months24 months | GAD-7, K-10, MINI, MINI-SPIN, NEO-FFI-N, PDSS-SR, PHQ-9, SDS |
| Hazen 2009 | USA | University students | PSWQ>60 | 24 | 23 | 3-6 weeks | PSWQ, BDI, STAI-T |
| Hirsch 2018 | UK | General - adults | Mixed sample anxiety/depression**#**GAD diagnosis on SCID-IGAD7>10 | 64 | 64 | 3-4 weeks1 month | GAD-7, PSWQ, PHQ-9, RRS  |
| Howell 2018 | USA | University students | Mixed sample non-clinical (GAD<4) and clinical mild GAD (4<GAD7<10)**~** | 197 | NR | 3 months | GAD-7**◊** |
| Johansson 2013 | Sweden | General - adults | Mixed sample anxiety/depression**#**GAD diagnosis on MINI GAD7>10  | 43 | NR | 10 weeks3 months | GAD-7, PHQ-9  |
| Jones 2016 | Canada | Over 60s | GAD diagnosis or threshold sub-clinical on MINIGAD7>10  | 46 | 41 | 7-10 weeks1 month | GAD-7, ACES, GAI, GDS, PHQ-9, PSWQ-A, WHOQOL |
| Navarro-Haro 2019 | Spain | Primary care - adults | GAD diagnosis on MINI | 42 | 30 | 7-12 weeks | GAD-7, DERS, FFMQ, HADS, MAIA |
| Paxling 2011 | Sweden | General - adults | GAD diagnosis on SCID-I PSWQ >53GADQ-IV>5.7 | 89 | 72 | 8 weeks1 year\*3 years\* | PSWQ, BAI, BDI-II, GADQ-IV, MADRS-S, QOLI, STAI-S, STAI-T |
| Pham 2016 | UK | General - adults | Mixed sample common mental health problems**##** GAD7 >6OASIS>8ASI>16 | 63 | 42 | 4 weeks | GAD-7, Acceptability, ASI-3, OASIS, PDSS-SR, QLES-Q-SF |
| Repetto 2013 (linked with (Pallavicini 2009, Gorini 2010a, 2010b) | Italy | Primary care - adults | GAD diagnosis (unspecified tool used) | 25 | 24 | Not reported | GAD-7, PSWQ, BAI, HAM-A, STAI |
| Richards 2016 | Ireland | University students | GAD7>10 | 137 | 112 | 6 weeks | GAD-7, PSWQ, BDI-II, WASA |
| Robinson 2010 | Australia | General - adults | GAD diagnosis on MINI | 150 | 138 | 11 weeks3 months\* | GAD-7, PSWQ, K-10, PHQ-9, SDS |
| Teng 2019 | Taiwan | General - adults | GAD diagnosis on DIS-IVPSWQ >60 | 93 | 82 | 4 weeks1 month | PSWQ, BAI, BDI, STAI-S, STAI-T |
| Titov 2009 (linked with Lorian 2012) | Australia | General - adults | GAD diagnosis on MINI | 34 | NR | 9 weeks | GAD-7, K-10, PHQ-9, SDS |
| Titov 2010 | Australia | General - adults | Mixed sample anxiety/depression**#** GAD diagnosis on MINI | 48 | 19 | 9 weeks3 months | PSWQ, DASS-21, K-10, NEO-FFI-N, PDSS-SR, PHQ-9, SDS, SPSQ |
| Topper 2017 | Netherlands  | 15 to 22-year-olds | PSWQ above 66th percentile (score 47)  | 251 | 218 | 8-10 weeks3 months12 months | PSWQ, BDI-II, EDI-2-BU, GADQ-IV, MASQ-D30, PTQ, QDS, RRS |

\* Patients received active intervention at follow-up timepoint.

◊ Study not included in the meta-analysis because it reported only categorical GAD-7 outcomes.

# Outcomes reported separately for GAD sub-group

## Outcomes reported for the whole mixed sample of common mental health problems but all participants scored GAD7>6

~ Outcomes reported separately for mild GAD sub-group (4<GAD7<10)

++ For the full measure names corresponding to the acronyms, see appendix E.

**Table 2 – Characteristics and classifications of interventions and controls in RCTs of digital interventions for GAD**

| **Study**  | **Intervention or Control Description**(mode of delivery, therapy/control method, type of interpersonal contact/support) | **Classification** |
| --- | --- | --- |
| Andersson 2012 | 1. Web-based psychodynamic therapy + weekly online support by psychology students/qualified psychologist
 | SDI  |
|  | 1. Web-based CBT + weekly online support by psychology students/qualified psychologists
 | SDI |
|  | 1. Waiting list (crossover to web-based CBT at 3-months).
 | NI |
| Andersson 2016 | 1. Web-based extinction therapy + daily online support by psychology students
 | SDI |
|  | 1. Waiting list (+ weekly online# PSWQ ratings and option to phone if symptoms worsen - crossover to web-based extinction therapy at 10 weeks)
 | NI1 |
| Christensen 2014a | 1. Web-based CBT - no interpersonal communication
 | UDI |
|  | 1. Web-based CBT + weekly phone calls by "casual interviewers"
 | SDI |
|  | 1. Web-based CBT + weekly reminder email similar in content to phone calls by “casual interviewers” but no two-way communication
 | UDI |
|  | 1. Control website (information about general health) - no interpersonal communication
 | UDC |
|  | 1. Control website (information general health) + weekly phone calls by "casual interviewers"
 | SDC |
| Christensen 2014b | 1. Web-based CBT + scheduled on-site meetings with psychologists/GPs
 | SDI |
|  | 1. Control website (information about general health) + scheduled meetings with psychologists/GPs
 | SDC |
|  | 1. Medication (SSRI – Sertraline 25mg up to 100mg per day) + scheduled meetings with psychologists/GPs
 | M |
| Dahlin 2016 | 1. Web-based MBT & ACT + weekly messages via a secure messaging system by psychology students
 | SDI |
|  | 1. Waiting list (+weekly online# GAD-7 and PSWQ ratings – contact with administrator implied for weekly measure completion but unclear – crossover to modified web-based MBT & ACT at 9 weeks)
 | NI1 |
| Dear 2015\* | 1. Web-based CBT (trans-diagnostic model focusing on mental wellbeing) + weekly phone/email contact with qualified psychologists
 | SDI |
|  | 1. Web-based CBT (GAD-specific focusing on worry control) + weekly phone/email contact with qualified psychologists
 | SDI |
|  | 1. Web-based CBT (trans-diagnostic model focusing on mental wellbeing) + standardised weekly email reminders and option to phone/email if needed technical support or had other problems - no scheduled or regular interpersonal contact
 | UDI |
|  | 1. Web-based CBT (GAD-specific focusing on worry control) + standardised weekly email reminders and option to phone/email if needed technical support or had other problems - no scheduled or regular interpersonal contact
 | UDI |
| Hazen 2009 | 1. Computer-delivered attentional retraining + ‘non-therapy’ meetings with “experimenters” every 6 days
 | SDI |
|  | 1. Sham training + ‘non-therapy’ meetings with “experimenters” every 6 days
 | SDC |
| Hirsch 2018\*\* | 1. Web-based CBM + 1 initial onsite meeting + regular (unspecified) contact by phone/email/SMS with researchers (unspecified qualifications) + RNT priming
 | SDI |
|  | 1. Web-based CBM + 1 initial onsite meeting + regular (unspecified) contact by phone/email/SMS with researchers (unspecified qualifications) – no RNT priming
 | SDI |
|  | 1. Control website (neutral scenarios) + 1 initial onsite meeting + regular (unspecified) contact by phone/email/SMS with researchers (unspecified qualifications)
 | SDC |
| Johansson 2013 | 1. Web-based psychodynamic therapy + weekly written messages via online messaging system by therapists (unspecified qualifications)
 | SDI |
|  | 1. Waiting list + weekly assessment and non-directive support via online messaging system with therapists (unspecified qualifications) matching therapist support in the intervention
 | SDC2 |
| Jones 2016 | 1. Web-based CBT + weekly messages via online messaging system by therapists (unspecified qualification)
 | SDI |
|  | 1. Waiting list (crossover after 7-10 weeks) – no monitoring or other input specified
 | NI |
| Navarro-Haro 2019 | 1. Group MBI in weekly onsite meetings with a therapist
 | SNoDI |
|  | 1. VR mindfulness skills + group MBI in weekly onsite meetings with a therapist
 | SDI |
| Paxling 2011 | 1. Web-based CBT (like an online book) + weekly online/email contact with therapist
 | SDI |
|  | 1. Waiting list (crossover after 8 weeks) - no monitoring or other input specified
 | NI |
| Pham 2016 | 1. Mobile game of breathing re-training – no interpersonal contact
 | UDI |
|  | 1. Waiting list + weekly newsletter with curated content on breathing retraining exercises, matching content to mobile game, mindfulness meditation (assumed via mobile but not clear) + email reminders to complete assessments (crossover to access the mobile game after 4 weeks)
 | UDC3 |
| Repetto 2013  | 1. VR relaxation during weekly meetings with therapist + mobile phone home-access of VR environments
 | SDI |
|  | 1. VR relaxation during weekly meetings with therapist + mobile phone home-access ofVR environments + biofeedback machine for therapist to adapt VR environments according to participant heart rate
 | SDI |
|  | 1. Waiting list (no monitoring or any other input specified)
 | NI |
| Richards 2016  | 1. Web-based CBT + weekly online messages by psychologists
 | SDI |
|  | 1. Waiting list (crossover at week 7 – no monitoring or other input)
 | NI |
| Robinson 2010  | 1. Web-based CBT + weekly phone/email contact by a “clinician” (clinical psychologist)
 | SDI  |
|  | 1. Web-based CBT + weekly phone/email contact by a “technician” (administrative clinic manager)
 | SDI  |
|  | 1. Waiting list (crossover at week 11 - no monitoring or other input)
 | NI |
| Teng 2019  | 1. Mobile app - home-delivered ABM + weekly ‘lab’ meeting with assistant + phone call if missed sessions
 | SDI |
|  | 1. Mobile app - attention training + weekly ‘lab’ meeting with assistant + phone call from if missed sessions
 | SDC |
|  | 1. Waiting list + weekly meetings with research assistant for matching assessment in a ‘lab’
 | SNoDC4 |
| Titov 2009 | 1. Web-based CBT + moderated online discussion forum + instant online messaging + 1 initial phone contact + subsequent email/phone weekly contact with clinical psychologist
 | SDI |
|  | 1. Waiting list (crossover at week 11 – no monitoring or other input)
 | NI |
| Titov 2010  | 1. Web-based CBT + moderated online discussion forum + instant online messaging + 1 initial phone contact + subsequent email/phone weekly contact with clinical psychologist
 | SDI |
|  | 1. Waiting list (+ unclear if contact with psychologist - crossover at week 9)
 | NI |
| Topper 2017 | 1. Group CBT in weekly meetings with psychologists
 | SNoDI |
|  | 1. Web-based CBT + weekly online personalised feedback from psychologists (unclear whether there was two-way communication between participant and psychologist in response to feedback).
 | SDI |
|  | 1. Waiting list (crossover at 12 months – no monitoring or other input from the research team)
 | NI |

# not clear whether online measures were sent by email or were automated and delivered via a platform

\* 2x2 factorial RCT, so outcomes reported in 2 groups (SDI vs. UDI for both GAD-specific and transdiagnostic groups)

\*\* Outcomes reported as a single group across both interventions (SDI with or without RNT)

\*\*\* Study not included in the meta-analysis because it only reported GAD-7 categorical outcomes

1 ‘Waiting list’ classified as ‘No Intervention’ (NI) because the online weekly ratings were self-completed without any further input from the research team

2 ‘Waiting list’ classified as ‘Supported Digital Control’ (SDC) because there was substantial and regular non-specific support and monitoring via an online messaging system with a therapist that matched the duration of therapeutic support in the intervention arm

3 ‘Waiting list’ classified as Unsupported Digital Control (UDC) because there was substantial and regular information and therapeutic advice similar to that of the intervention –via the mobile (inferred) (hence ‘digital’); standardised materials without two-way interaction with a therapist (hence ‘unsupported’)

4 ‘Waiting list’ classified as Supported Non-Digital Control (SNoDC) because there was substantial and regular face-to-face assessments with a researcher that matched the assessments of the intervention groups.

ABM = Attentional Bias Modification; ACT = Acceptance and Commitment Therapy; CBT = Cognitive Behaviour Therapy; GAD=Generalised Anxiety Disorder; M=medication; MBI = Mindfulness Based Intervention; MBT = Mindfulness Based Therapy; NI=No intervention; PSWQ=Penn State Worry Questionnaire; RNT = Repetitive Negative Thinking; SDC=Supported Digital Control; SDI=Supported Digital Intervention; SNoDC=Supported Non-Digital Control; SNoDI=Supported Non-digital intervention; SSRI = Selective Serotonin Reuptake Inhibitor; VR = Virtual Reality; UDC=Unsupported Digital Control; UDI=Unsupported Digital Intervention.

**Table 3 - Full meta-analysis results: network and direct pairwise comparisons between all interventions and controls for post-treatment (12 weeks) GAD-7 scores adjusted for baseline**

|  |  |  |
| --- | --- | --- |
|   |   | **GAD-7 Direct Pairwise\*,\*\*** |
|   | **Comparator** | **NI** | **UDC** | **SDC** | **UDI** | **SDI** | **SNoDI** | **M** |
|   | **FE** | **RE** |
| **GAD-7 Network**  | **NI** |   | Not available  | Not available  | Not available  | -3.65[-8.19, 0.9] | -1.26\*\*[-43.8, 40.93] | Not available  | Not available  |
| **UDC** | -0.26[-16.82, 16.42] |  | -0.80\*[-12.30, 10.70] | -0.77[-7.81, 6.26] | -1.50\*[-11.34, 8.34] |  Not available | Not available  |
| **SDC** | -0.14[-15.52, 15.30] | 0.12[-9.84, 10.05] |   | -0.51[-31.75, 28.95] | -8.16[-26.57, 13.04] | -1.07\*\*[-30.2, 27.19] | Not available  | -8.20\*[-22.07, 5.67] |
| **UDI** | -0.96[-16.43, 14.54] | -0.7[-9.10, 7.58] | -0.83[-8.82, 7.29] |  ` | -0.14[-8.85, 8.89] | -1.38\*\*[-26.23, 23.62] | Not available  | Not available  |
| **SDI** | -1.75[-15.72, 12.26] | -1.49[-10.51, 7.47] | -1.61[-8.03, 4.82] | -0.79[-7.39, 5.80] |  | -0.71\*[-14.03, 12.61] | -2.70\*[-9.80, 4.40] |
| **SNoDI** | -2.41[-22.34, 17.46] | -2.21[-18.69, 14.5] | -2.28[-17.66, 13.16] | -1.45[-16.92, 14.03] | --- | -0.65[-14.67, 13.23] |   | Not available  |
| **M** | -4.95[-21.09, 11.26] | -4.69[-16.62, 7.11] | -4.81[-14.59, 4.83] | -3.97[-14.3, 6.20] | --- | -3.18[-11.32, 4.76] | -2.56[-18.71, 13.54] |   |

FE=Fixed Effects; GAD=Generalised Anxiety Disorder; M=medication; NI=No intervention; RE=Random Effects; SDC=Supported Digital Control; SDI=Supported Digital Intervention; SNoDI=Supported Non-digital intervention; UDC=Unsupported Digital Control; UDI=Unsupported Digital Intervention.

\*Non-pooled data for when n=1.

\*\*Pairwise ANCOVA RE meta-analysis for when n>3 for contrasts with intervention SDI.

Lower left triangle: Network Meta-Analysis Results (ANCOVA RE); Upper right triangle (shaded area): Direct Pairwise Meta-Analysis Results (ANCOVA FE); Cells with thick black perimeter: available comparisons between digital interventions (supported and unsupported) and alternatives

# Figures

**Fig. 1- PRISMA diagram for the identification and selection of clinical trials relating to digital interventions for generalised anxiety disorder (GAD)**

Adapted from: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

**Fig. 2- Network plot for comparisons between all interventions and controls for GAD populations in RCTs with GAD-7 score**

GAD=Generalised Anxiety Disorder; M=medication; NI=No intervention; SDC=Supported Digital Control; SDI=Supported Digital Intervention; SNoDI=Supported Non-digital intervention; UDC=Unsupported Digital Control; UDI=Unsupported Digital Intervention.

* Line thickness around the node: proportional to the number of patients contributing to each intervention/control group
* Line thickness connecting nodes: proportional to the number of patients contributing to each pairwise comparison between interventions/controls
* n= number of trials informing each comparison