Journal of Affective Disorders 369 (2025) 1031–1044

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Review article

Digital mental health interventions for the treatment of depression: A multiverse meta-analysis

Constantin Yves Plessen [a,b,*,1], Olga Maria Panagiotopoulou [a,1], Lingyao Tong [a], Pim Cuijpers [a], Eirini Karyotaki [a]

[a] *Department of Clinical, Neuro-, and Developmental Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands*
[b] *Charité University Medicine Berlin, Berlin, Germany*

ARTICLE INFO

Keywords:
Meta-analysis
Depression
Digital interventions
Multiverse meta-analysis
Vibration of effects

ABSTRACT

Background: The varying sizes of effects in published meta-analyses on digital interventions for depression prompt questions about their efficacy.
Methods: A systematic search in Embase, PsycINFO, and PubMed identified 125 randomised controlled trials up to February 2023, comparing digital interventions for depression against inactive controls. The stability of results was evaluated with a multiverse meta-analysis, thousands of meta-analyses were conducted based on different combinations of analytical choices, like target populations, intervention characteristics, and study designs.
Results: A total of 3638 meta-analyses were performed based on 125 randomised controlled trials and 263 effect sizes, with a total of 32,733 participants. The average effect size was Hedges' $g = 0.43$, remaining positive at both the 10th ($g = 0.16$) and 90th percentiles ($g = 0.74$). Most meta-analyses indicated a statistically significant benefit of digital interventions. Larger effects were observed in meta-analyses focusing on adults, low- and middle-income countries, guided interventions, comparing interventions with waitlist controls, and patients with major depressive or unipolar mood disorders. Smaller effects appeared when adjusting for publication bias and in assessments after 24 weeks.
Limitations: While multiverse meta-analysis aims to exhaustively investigate various analytical decisions, some subjectivity remains due to the necessity of making choices that affect the methodology. Additionally, the quality of the included primary studies was low.
Conclusions: The analytical decisions made during performing pairwise meta-analyses result in vibrations from small to medium effect sizes. Our study provides robust evidence for the effectiveness of digital interventions for depression while highlighting important factors associated with treatment outcomes.

1. Introduction

Depression is often associated with important personal challenges, significant productivity loss and economic costs, and in certain cases, reduced life expectancy (Evans-Lacko and Knapp, 2016; Fried and Nesse, 2014; Jain et al., 2022; Laursen et al., 2016; Stewart, 2003). Despite the joint efforts of clinicians and researchers to alleviate the burden of depression, it remains highly prevalent among the global population with severe consequences both for the individual and the society as a whole (Johnston et al., 2019; Santomauro et al., 2021; World Health Organization, 2023). International guidelines promote evidence-based psychotherapy and/or pharmacotherapy as first-line treatments, depending on many factors such as the individual's symptom severity (American Psychological Association, 2019). However, even though effective, access to traditional psychotherapy is limited for multiple reasons including high costs, shortage of personnel and self-stigmatisation (Carbonell et al., 2020; Ebert et al., 2017; Mohr et al., 2006; Schaffler et al., 2022). As a consequence, most people do not receive adequate or any form of treatment at all (Moitra et al., 2022).

Over the past years, technology has transformed how we perceive mental health care by allowing us to provide evidence-based treatments, such as cognitive behavioural therapy, in a cost-effective manner. Digital interventions, namely guided and self-guided online programs or smartphone apps based on different psychological treatments, offer

* Corresponding author at: Department of Clinical Psychology, VU University Amsterdam, Van der Boechorststraat 7, 1081 BT Amsterdam, the Netherlands.
E-mail address: yves.plessen@proton.me (C.Y. Plessen).
[1] Contributed equally and share first authorship.

https://doi.org/10.1016/j.jad.2024.10.018
Received 23 December 2023; Received in revised form 2 October 2024; Accepted 7 October 2024
Available online 15 October 2024

many innovative and practical solutions such as anonymity, time flexibility, adaptability and scalability, among others (Cuijpers and Riper, 2015; Patel et al., 2020; Rodriguez-Villa et al., 2020). Due to the widespread implementation of internet-based interventions for depression, numerous randomised controlled trials on individuals with depression have been conducted all over the world, and thus, many meta-analyses have examined the overall effectiveness of such interventions in reducing the symptoms of depression. Evidence suggests that digital interventions can be effective in facilitating the needs of people facing depression by significantly reducing clinical symptoms by the end of the treatment period (Josephine et al., 2017; Karyotaki et al., 2021; Lindegaard et al., 2020; Roman et al., 2020; Sierra et al., 2018; Yang et al., 2018) while sometimes this decline has shown to persist in the long-term (Köhnen et al., 2021; Reins et al., 2021; Sztein et al., 2018). Most importantly, recent findings have highlighted that digital and face-to-face interventions show comparable treatment effects (Ahern et al., 2018; Cuijpers et al., 2019; Hedman-Lagerlöf et al., 2023; Higinbotham et al., 2020; Kheirkhah et al., 2023; Moshe et al., 2021). Additionally, research indicates that individuals experiencing moderate to severe symptoms of depression tend to benefit more from digital interventions than those experiencing less severe symptoms, underscoring the role of baseline severity as a strong predictor of treatment response (Chan et al., 2022; Furukawa et al., 2021; Karyotaki et al., 2021; Serrano-Ripoll et al., 2022).

Although a considerable number of meta-analyses exploring the effectiveness of digital interventions for depression have been published in recent years, their findings regarding effect sizes are inconsistent, ranging from small to large (Hedges' g 0.22 to 1.01). These variations can be attributed to numerous factors, including differences in populations, interventions, comparators, outcomes, as well as the distinct study characteristics and statistical models employed in the meta-analyses (Chan et al., 2022; Cuijpers et al., 2019; Firth et al., 2017; Han and Kim, 2022; Josephine et al., 2017; Karyotaki et al., 2017; Köhnen et al., 2021; Pang et al., 2021; Serrano-Ripoll et al., 2022; Sierra et al., 2018; Xiong et al., 2023). The broad range of effect size estimates highlights potential issues with the robustness of conclusions about the effectiveness of online interventions. Variability in methodological approaches and the inherent subjectivity in the selection of data and meta-analytic techniques may have influenced these discrepancies, underscoring the necessity for meticulous examination of analytical decisions prior to establishing firm conclusions (Simmons et al., 2011; Voracek et al., 2019).

In light of the substantial body of meta-analyses evaluating the effectiveness of digital interventions for depression, it becomes imperative to investigate the robustness of this evidence to better understand its strengths and limitations. This will not only aid in clinical advancement but also support the ongoing development of the field. Consequently, our study aims to conduct a multiverse meta-analysis—a comprehensive sensitivity analysis—to scrutinise the robustness of meta-analytic findings related to digital interventions for depression. Compared to traditional meta-analyses, where researchers make a single decision on the most appropriate statistical analysis and which specific studies to include, a multiverse meta-analysis aims to explore many available methodological paths and combinations of studies (Voracek et al., 2019). This becomes possible by including a broad range of studies and performing multiple meta-analyses, each time pooling a different subgroup of studies (Which factors) while using various statistical approaches (How Factors). This exhaustive investigation of the existing literature expands our current knowledge while also establishing the robustness of previous findings. More specifically, a multiverse meta-analysis enables us: (1) statistically integrate all existing meta-analyses; (2) explore all conceivable and justifiable meta-analyses derivable from current data that have not yet been conducted; (3) assess the impact of different analytical choices on outcomes; (4) identify literature gaps and investigating the reasons behind divergent findings; and (5) visualise the emerging results to provide a comprehensive overview of the field (Simonsohn et al., 2020; Steegen et al., 2016).

Thus, the primary research question of our investigation is to determine whether the majority of meta-analyses substantiate the effectiveness of digital interventions in reducing depressive symptoms. Furthermore, this study explores how the outcomes vary based on different target populations, intervention characteristics including the level of support, technological format, psychotherapeutic approaches, control conditions, measured outcomes, and study design factors such as the risk of bias and measurement timing. Thereby, we examine the variations in summary effect sizes across the multiverse of all possible meta-analytic configurations.

2. Methods

2.1. Protocol

The present study was not preregistered due to the exploratory and innovative nature of the multiverse meta-analysis methodology, which allows for a more flexible approach by incorporating and assessing various analytical decisions. More specifically, we evaluated all possible combinations of possible factors according to our inclusion criteria (PICO) and employed most non-arbitrary, well-established meta-analytic approaches. The multiverse analysis inherently functions as a sensitivity analysis since all analyses are reported, which mitigates the risks associated with a non-registered analysis. To ensure transparency and reproducibility, we have thoroughly reported our analytical decisions. All components necessary for reproducible data analysis (open data, open code) were made accessible via the OSF (https://osf.io/zagvt/?view_only=65adeec8b8784774841d8b28c551776f) and comply with the FAIR (findable, accessible, interoperable, reusable) guiding principles for scientific data (Wilkinson et al., 2016).

2.2. Eligibility

Randomised controlled trials were deemed eligible if they examined the effectiveness of digital interventions compared to an inactive control for the treatment of depression. The following inclusion criteria were assessed for each primary study: a) Participants of all ages had to be recruited based on elevated depression symptoms as indicated by a validated self-reported questionnaire (suggesting at least mild depression) or a clinical diagnosis of any depressive disorder according to a structured diagnostic interview at the time of enrolment, b) Digital guided or self-guided interventions founded on therapeutic protocols focusing on the treatment of depression (cognitive behavioural activation, behavioural activation, problem-solving therapy, etc.). These interventions had to be delivered through mobile apps or internet-based platforms and had to be accessible remotely through the internet, c) Comparison groups had to be purely inactive or providing minimal psychoeducation (care-as-usual, waiting list, no treatment, attention control), d) Sufficient data to calculate effect sizes for depression levels as a primary outcome had to be reported. Language restrictions were not applied, and studies were included regardless of whether participants had physical or mental comorbidities.

Studies were excluded if: a) digital interventions were delivered as an adjunctive to face-to-face psychotherapy or in a blended format, b) the treatment was part of a dismantling study, a stepped-care program, or maintenance trials aimed at the prevention of relapse in previously depressed patients.

2.3. Search strategy and selection process

Relevant records were identified from an existing repository including all available studies on the effectiveness of digital interventions in reducing the symptoms of depression and/or anxiety (randomised controlled trials, observational studies conducted in primary care, systematic reviews and meta-analyses). The repository was established on February 25th, 2022, by conducting an extensive literature search on Embase, PsycINFO and PubMed, and it is updated yearly (Last update: February 6th, 2023) (Karyotaki et al., 2024). The extensive search was filtered for records published after 2000 and only peer-reviewed papers were examined for inclusion. All records were screened by title and abstract, and if judged as possibly eligible they were examined on a full-text basis by three pairs of independent researchers. Disagreement was solved through discussion. The full search strings are presented in the supplementary material (Text S1). Additional records were added to the repository from the meta-analytic database of psychological treatments for depression (Cuijpers et al., 2008) and through known references.

2.4. Data collection process

Data for each eligible study were extracted manually by pairs of independent reviewers. Discrepancies were solved through discussion and a senior researcher was consulted if necessary. From each study, we extracted data on the characteristics of the included studies (author details, year of publication, sample size in each group), demographic characteristics of participants (age, gender), and depression outcomes. We extracted all available depression outcomes (self-reported or clinician-rated) reported in each paper at all time points (baseline, post-treatment, follow-ups).

We investigated a range of analytical decisions that meta-analysts might use when examining the effectiveness of digital interventions based on specific criteria for study inclusion. These criteria, so-called Which factors (determining which data to meta-analyse), were shaped by the PICOS framework, which considers population, intervention, control comparison, outcomes, and study design. The “Which factors” covered demographic data, intervention details (like delivery method, treatment type, and support levels), control groups, diagnoses, and study design elements (such as bias risk and assessment timing) and income level of the country in which the study was conducted. A comprehensive list and description of all “Which factors” (e.g. age ranges for the Which factor population) used to examine every possible combination in our multiverse meta-analysis is available in the supplementary material (Text S2).

Study risk of bias assessment: The risk of bias was assessed with the use of the revised Cochrane Collaboration's Risk of Bias tool 2.0 for individually randomised parallel-group trials (Sterne et al., 2019). The instrument follows five main domains with several subitems that can result in an overall score of low risk, some concerns, or high risk. The domains include: 1) Bias due to inconsistencies in the randomisation procedures (sequence generation, allocation concealment, sample imbalances), 2) Bias due to the deviations from intended interventions for the effect of the assignment to the intervention, 3) Bias introduced by unavailable outcome data, 4) Bias due to the outcome measurements incorporated in each study, 5) Bias due to selective outcome publication. Studies were rated as an overall low-risk when all five domains received a low-risk assessment. Studies were categorised as an overall high risk when they were rated as high risk in at least one domain or assessed as some concerns in at least three domains. In all other cases, studies were judged as an overall rating of some concerns.

2.5. Multiverse meta-analysis

Researchers have to decide between several equally defensible choices at multiple stages when conducting a meta-analysis: which studies to meta-analyse based on prespecified study inclusion criteria (*Which* factors) and how to analyse them (*How* factors), which involves deciding between different available effect size estimators. In a multiverse meta-analysis, researchers identify all possible stages of analytical decisions, determine reasonable choices for each stage, and implement all of them. When considering reasonable choices for the *How* factors, we included eight different meta-analytical models for pooling the summary effect size to provide a broad perspective: random effects, fixed effect, three-level, robust variance estimation (RVE), precision-effect test and precision-effect estimate with standard errors (PET-PEESE), *p*-uniform, Unrestricted Weighted Least Squares (UWLS), and Weighted Average of Adequately Powered models (WAAP).

We employed this range of advanced meta-analytic methods to expand the evidence base beyond standard approaches, exploring how different analytical choices influence the resulting summary effect sizes. While random effects models are commonly used to account for between-study variation (Harrer et al., 2021), additional methods like PET-PEESE and p-uniform specifically address publication bias, which random effects models alone do not correct for. PET-PEESE corrects bias by examining the relationship between effect size and standard error (Stanley et al., 2022), while p-uniform utilises the distribution of *p*-values to identify selection bias (Van Aert and Van Assen, 2018), adding layers of bias control that enhance the validity of our results.

Furthermore, three-level models and the RVE model address the issue of effect size dependency within studies, which is not adequately handled by standard random effects models. These methods account for both within-study and between-study variability, offering a more nuanced analysis of hierarchical data structures commonly encountered in meta-analytic research.

The Unrestricted Weighted Least Squares (UWLS) estimator, along with the Weighted Average of Adequately Powered (WAAP) estimator downweigh the influence of small studies and might have advantages over other estimators in meta-analyses. Simulation studies have shown that these advantages prevail even when sample sizes and heterogeneity vary—and both in the presence and absence of publication bias (Stanley et al., 2022). While UWLS yields the same point estimate as the fixed effect model, the standard errors and confidence intervals are larger when there is more heterogeneity (Stanley and Doucouliagos, 2014). WAAP excludes small studies and only includes studies with 80 % or higher statistical power. Simulation studies suggest that when there is publication-selection bias, WAAP is less biased than other weighted average estimators, such as the random-effects, fixed-effects, and UWLS estimators. For an in-depth explanation and the reasoning behind including the first six methods, please see Plessen et al. (2023).

As a result, our multiverse meta-analysis could theoretically report a total of 233,280 meta-analyses resulting from all possible combinations of *Which* and *How* Factors. We decided to include only meta-analyses with data from at least 10 primary studies in our main analyses, as we aimed to provide valid evidence and some methods such as PET-PEESE require this amount of effect sizes to reach statistical power to.be interpretable (Stanley, 2017). We additionally conducted a sensitivity analysis with different cut-offs for the required meta-analysis size, namely using 2 studies, 5 studies, 25 studies, and 50 studies accordingly. These can be found in the Online Supplemental material.

2.5.1. Descriptive specification curve

We used descriptive specification curve plots to inspect gaps and patterns in the meta-analytic summary effects for all *Which* and *How* Factor combinations (Voracek et al., 2019). The descriptive meta-analytic specification plot displays all possible meta-analyses and visualises each specification's *Which* and *How* Factor combination, including the resulting meta-analytic summary effects ordered by magnitude with their respective 95 % confidence intervals.

We reported the percentages of meta-analyses that produced a) summary effect sizes, and b) 95 % confidence intervals larger than two relevant cut-offs: either a null effect (Hedges' $g = 0$) or a clinically relevant effect size of Hedges' $g = 0.24$ (Cuijpers et al., 2014).

2.5.2. Vibrations of effects

The Vibration of effects (VoE) is a graphical method for analysing the variability in meta-analysis outcomes. It visually represents the range of possible results based on different analytical decisions (Patel et al., 2015). In the VoE plot, the *p*-values are plotted on the vertical (y) axis, and effect sizes (ES) are on the horizontal (x) axis. Each illustrated point corresponds to the results of a specific meta-analysis. This plot shows the range of effect sizes across various meta-analyses, their statistical significance (notably at $p < .05$; El Bahri et al., 2022), and their frequency, with warmer colours indicating higher occurrence of meta-analyses and cooler colours indicating fewer. The VoE aims to highlight the robustness of the evidence and examine how certain analytical choices, like the types of interventions, control groups, and statistical methods, impact meta-analysis results. A smaller vibration of effects suggests more certainty in the observed results, whereas a larger vibration of effects serves as an indicator to further examine the reliability of the findings and assess the impact of certain analytical choices. Similarly, a heat map was drawn to show VoE for heterogeneity. Higgins I^2 was plotted on the x-axis, and the logarithm of the *p*-value of the *Q*-test on the y-axis.

An important aspect we considered is the identification of a substantial VoE, characterised as a Janus effect (Patel et al., 2015). This occurs when the direction of an effect differs at opposite ends of the distribution, specifically comparing the 90th percentile (Hedges' $g > 0$) with the 10th percentile (Hedges' $g < 0$). This concept, expanded upon by El Bahri et al. (2022), helps in understanding the impact of different analytical choices on the outcomes of meta-analyses.

2.5.3. Exploration of analytical decisions associated with the magnitude of effect sizes

To ensure robust results and mitigate potential distortions from non-normally distributed magnitudes of effect sizes, we examined the relationship between each methodological choice and the variation in effect size using multiple median regressions. Additionally, we created descriptive specification curve plots (Voracek et al., 2019) and rain cloud plots (Allen et al., 2021) to visually inspect the differences between each analytic decision. Raincloud plots blend the aspects of box plots, violin plots, and scatter plots into a single graphic. The box plot element highlights the data's interquartile range, giving insights into the spread and central tendency of the data. The violin plot, forming the "cloud," visually represents the density of the data, indicating areas of higher concentration. Lastly, the scatter plot, or the "rain" part, features individual data points depicted as dots beneath the violin plot's "cloud," resembling raindrops. This combination in a raincloud plot provides a multifaceted view of the data, encompassing distribution, density, and individual data points.

2.5.4. Statistical models

All statistical analyses were carried out with the use of the Metafor (Viechtbauer, 2010) and puniform* packages, in R (version 4.2.2). To investigate the effectiveness of digital interventions in reducing the symptoms of depression when compared to control, we used continuous data (means, standard deviations, number of participants in each group) and computed the standardised mean differences between groups at post-treatment and follow-up assessments (expressed as Hedges' *g*). If these data were not reported, we retrieved alternative statistical information (predefined order of data extraction: dichotomous outcomes, mean change, *p* values) that were converted to effect sizes with the use of the metapsyTools package (Harrer et al., 2022).

3. Results

3.1. Study selection

The database search in Pubmed, Embase and PsycINFO yielded 19,579 records. After the removal of 8749 duplicates, 10,830 titles and abstracts were screened. Subsequently, 1151 records, 6 of which were unretrievable, were assessed on a full-text basis, leading to the inclusion of 112 eligible papers. Additionally, 9 more papers were included from an existing depression database (Cuijpers et al., 2022) and 3 from previously known references resulting in a total of 124 eligible papers for the multiverse meta-analysis, one of which reported two independent randomised controlled trials. These 125 trials contained 263 effect sizes with a total of 32,733 participants.

For a full description of the screening process, please consult the PRISMA Flowchart (Fig. 1.). Table 1 presents the summary characteristics of all included primary studies. See Table S1 and S2 for the summary characteristics of the included effect sizes.

3.2. Multiverse meta-analysis

To ensure the adequate and precise calculation of summary effects, we included meta-analyses with at least 10 primary studies each in our main analyses, resulting in a total of 3638 meta-analyses based on all possible combinations of our specified *Which* and *How* factors.

Among the 3638 conducted meta-analyses, effect sizes for the effectiveness of digital interventions for depression ranged from Hedges' $g = -0.34$ to 1.42, with a median of 0.43. The effect sizes were positive at both the 10th percentile (Hedges' $g = 0.16$) and the 90th percentile (Hedges' $g = 0.74$). Approximately 9 % (n = 328) of these meta-analyses included 10 trials, 27 % (n = 989) >25 primary studies, and 8 % (n = 283) >50.

For example, one meta-analysis using a three-level model encompassing all effect sizes from all primary studies yielded an effect size of $g = 0.58$ (95 % CI: 0.38–0.77; $p < .001$). The estimated variance components were $\tau^2_{\text{Level 3}} = 0.1615$ and $\tau^2_{\text{Level 2}} = 0.00$, indicating that overall heterogeneity is high, and that 86.6 % ($I^2_{\text{Level 3}}$) of the total variation is due to between-study heterogeneity, while 0 % ($I^2_{\text{Level 2}}$) is due to within-study heterogeneity.

3.2.1. Descriptive specification curve and vibration of effects

The descriptive specification curve showed that the summary effect sizes of meta-analyses can vary from null effects to large effect sizes. See Fig. 2 for a detailed visualization of effect sizes and their respective 95 % confidence intervals of the 3638 meta-analyses based on all possible combinations of *Which* and *How* factors. The estimated mean Hedges' *g* of the various subsets ranged from −0.34 to 1.42, with an interquartile range of 0.26 to 0.60. In total, 97 % of the estimated means were >0, and 82 % of these had 95 % *CIs* that did not include 0 (i.e., estimated means >0 which would have returned a statistically significant result, with a two-tailed *p*-value<.05).

In total, 78 % reached a clinically relevant effect size of Hedges' $g > 0.24$, and 50 % of the summary effect sizes had 95 % *CIs* larger than this clinically relevant effect size.

Across all meta-analyses, 364 had a Hedges' *g* below the 10th percentile (0 % of them with significant differences in favour of the control group) and 364 had a Hedges' *g* above the 90th percentile (98 % with significant differences in favour of the intervention). Among all meta-analyses, 2978 (82 %) showed a statistically significant difference in favour of digital interventions, while 558 (15 %) failed to do so. No meta-analysis showed a statistically significant difference in favour of the control condition.

Fig. 3 presents the heat map of the VoE. Among the meta-analyses of at least 10 trials, the sign of the Hedges' *g* at the 10th percentile and at the 90th percentile was positive for both cases ($g = 0.16$ vs. $g = 0.74$), suggesting the absence of substantial vibration. This pattern persisted for meta-analyses of at least 2, 5, 25 and 50 trials, see Fig. S1 for these sensitivity analyses.

Heterogeneity was assessed in 1264 meta-analyses, each with at least ten trials, where Higgins I^2 values could be calculated (fixed effect model, random effects model, 3-level model). The estimated I^2 values ranged from 0 % to 99 %, with a median of 76 %, a 1st quartile of 63 %, and a 3rd quartile of 87 % (The results are displayed graphically in Fig. S2).

Taken together, these results point to the overall effectiveness of digital health interventions in treating depressive disorders across a majority of the different combinations of *Which* factors and *How* factors.

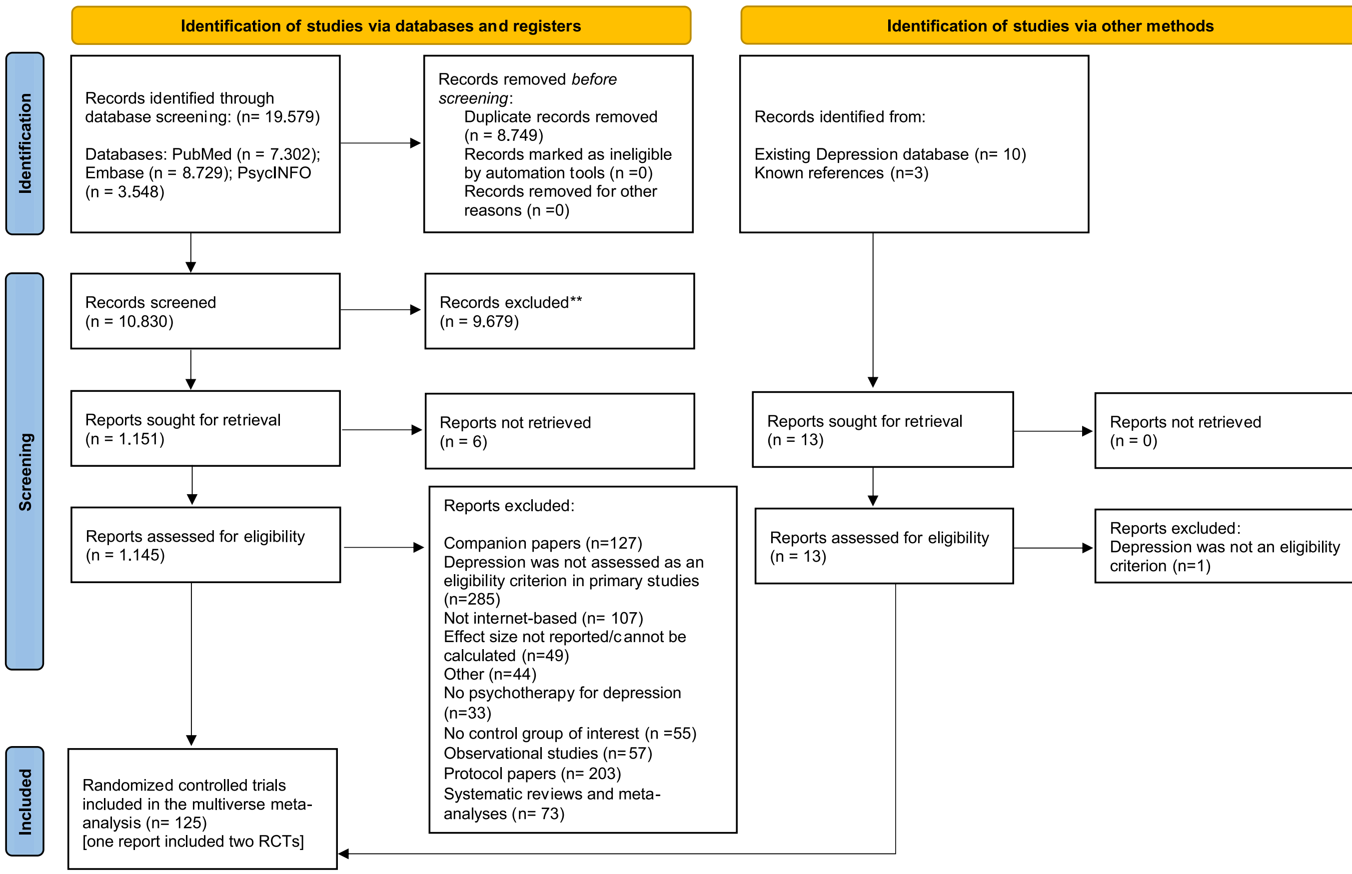

Fig. 1. PRISMA flowchart of the inclusion of primary studies.

We will describe specific patterns according to different *Which* and *How* factors in the following section.

3.3. Analytical decisions associated with VoE

The analytical decisions associated with the greatest increase in effect sizes (see Fig. 4) were the inclusion of studies using only guided interventions (+ 0.08 vs. inclusion of all types of guidance), comparison with waitlist control groups (+0.18 vs. inclusion of all control groups), the inclusion of mobile-based interventions (+0.10 vs. inclusion of all delivery technologies), the inclusion of studies restricted to low- and middle-income countries (+0.23 vs. all countries), and populations diagnosed with major depressive disorder (+0.12 vs. inclusion of all diagnoses).

The methodical choices associated with the greatest decreases were the comparison with care as usual as a control condition (−0.13 vs. all control conditions), the comparison at follow-up >24 weeks after the intervention (−0.19 vs. compared post-intervention), and the inclusion of other populations, like older adults or those not fitting into the remaining categories, (−0.15 vs. inclusion of all populations). Some populations were associated with lower effect size estimates (medical −0.10 and other groups −0.15 vs. inclusion of all populations). In terms of level of guidance, minimal to no support guidance was associated with lower effect size estimates (−0.09 vs. inclusion of all types of guidance).

Also, the use of four methods produced lower estimates, PET-PEESE (−0.25), fixed effect model (−0.08), WAAP (−0.12), UWLS (−0.08) compared with the RVE estimator.

Highlighted in red are the methodological choices that substantially impact the overall summary effect size, underscoring the critical nature of methodological decisions in meta-analytic outcomes.

Some *Which* factors appeared to consistently correlate with the magnitude of the summary effect, while others did not show a similar relationship. In the following section, these associations will be presented following their PICOs (population, intervention, control, outcomes, and study features). Effect sizes are presented as the median Hedges' g, whereas k signifies the number of produced meta-analyses for each subgroup.

3.3.1. Population

Target population. Population characteristics significantly influenced the summary effect sizes in the meta-analyses. Meta-analyses of the general adult population ($g = 0.50$, $k = 1022$) showed much larger effect sizes than those including participants with specific demographic characteristics, i.e. older adults, asylum seekers and refugees, migrants, employees in specific professions, veterans, ($g = 0.31$, $k = 30$). Meta-analyses encompassing all populations mirrored the adult population's results due to its larger sample size. Younger populations showed a slightly lower effect (Hedges' $g = 0.40$, $k = 10$), similar to studies focusing on perinatal ($g = 0.38$, $k = 67$), while medical groups had a marginally smaller effect ($g = 0.35$, $k = 240$). Overall, meta-analyses including all populations produced effect sizes similar to the adult population, as this was the largest proportion of meta-analyses ($g = 0.42$, $k = 2269$). Figs. 5 and 6 illustrate these variations in effect sizes by population type and a descriptive specification curve for adults, respectively.

Diagnosis. Summary effect sizes differed depending on the type of initial clinical assessment incorporated in the meta-analyses. Meta-analyses of participants diagnosed with Major Depressive Disorder yielded larger effects ($g = 0.61$, $k = 438$) compared to those using self-reported measures ($g = 0.40$, $k = 1140$) or focusing on other unipolar mood disorders ($g = 0.55$, $k = 67$). Figs. S10 and S11 elaborate on these results.

Table 1
Summary characteristics of included primary studies according to each possible methodological choice.

Summary characteristics of included primary studies.	
PICOS	
Characteristic	*N = 125*[a]
Population: Group	
Adults	66 (53 %)
Medical populations	21 (17 %)
Other populations	13 (10 %)
Women with perinatal depression	15 (12 %)
Young populations	10 (8.0 %)
Population: Diagnosis	
Diagnosis of major depressive disorder	26 (21 %)
Diagnosis of mood disorder	13 (10 %)
Self-reported questionnaire (exceeding a cut-off)	86 (69 %)
Intervention: Treatment modality	
CBT-based and not-CBT-based interventions	3 (2.4 %)
CBT-based interventions	96 (77 %)
Non-CBT-based interventions	26 (21 %)
Intervention: Format of delivery	
Mobile-based and website-based interventions	4 (3.2 %)
Mobile-based interventions	19 (15 %)
Website-based interventions	102 (82 %)
Intervention: Level of Support	
Compared multiple levels of support	9 (7.2 %)
Level 1: Minimal to no support	22 (18 %)
Level 2: Automated Encouragement	12 (9.6 %)
Level 3: Human Encouragement	13 (10 %)
Level 4: Guided	69 (55 %)
Control Condition	
Compared intervention to multiple control conditions	2 (1.6 %)
Compared to care as usual	32 (26 %)
Compared to other control conditions	27 (22 %)
Compared to a waitlist control condition	64 (51 %)
Study Design: Risk of Bias	
High Risk of Bias	40 (32 %)
Low Risk of Bias	17 (14 %)
Some Concerns	68 (54 %)
Study Design: Time Point	
Measures taken post-intervention	92 (74 %)
Measures taken at follow-up after 24 weeks	33 (26 %)
Study Design: Country	
High Income	111 (89 %)
Low- and Middle-Income	14 (11 %)

PICOS stands for: Patients; Intervention; Comparator; Outcomes; Study design.
[a] n (%).

See Fig. S3 for a raincloud plot and Fig. S4 for a descriptive specification curve plot focusing on MDD.

3.3.2. Intervention

Intervention modality. Whether the intervention was based on CBT principles or not produced negligible differences in effect size estimates. Non-CBT interventions had a slightly larger effect size (g = 0.46, k = 272) compared to the ones based on CBT ($g = 0.43$, $k = 1435$). However, only very few meta-analyses could exist based on purely non-CBT-based interventions (only 8 % of the produced meta-analyses included problem-solving therapy, psychodynamic therapy, life-review therapy, positive psychology, etc.). Similar results were observed in meta-analyses that included all treatment modalities ($g = 0.43$, k = 1931). See Fig. S5 for a raincloud plot and Fig. S6 for a descriptive specification curve plot focusing on non-CBT-based interventions.

Format of Delivery. We found similar summary effect sizes across different technologies used for the intervention delivery. Mobile-based interventions ($g = 0.49$, $k = 58$) showed slightly larger effects than internet-based (median $g = 0.44$, $k = 1593$). See Fig. S7 for a raincloud plot and Fig. S8 for a descriptive specification curve plot focusing on mobile-based interventions.

Level of support. Summary effect sizes differed depending on the level of support utilised in the intervention. Guided interventions ($g = 0.51$, $k = 1122$) had larger effect sizes than those with minimal or no support ($g = 0.37$, $k = 224$) and automated encouragement ($g = 0.32$, $k = 76$). Human encouragement interventions showed a marginally larger effect ($g = 0.39$, $k = 54$). See Fig. S9 for a raincloud plot and Fig. S10 for a descriptive specification curve plot focusing on self-guided interventions.

3.3.3. Comparison

Control Group. Summary effect sizes varied depending on the type of control group employed in the meta-analyses. Waiting list control groups resulted in larger effect sizes (g = 0.64, k = 872) compared to the small effects found when care as usual control groups were the comparator ($g = 0.25$, $k = 361$), or other types of control, such as attention control or minimal psychoeducation ($g = 0.32$, $k = 232$). See Fig. S11 for a raincloud plot and Fig. S12 for a descriptive specification curve plot focusing on care-as-usual control groups.

3.4. Study design

Strategies in Dealing with the Risk of Bias of Primary Studies. The choice in dealing with the risk of bias by either excluding high-risk of bias studies, solely including low risk of bias studies, or incorporating all studies, resulted in significantly different effect size estimates. Including all studies ($g = 0.46$, $k = 2192$) showed larger effects than excluding high-risk studies ($g = 0.40$, $k = 1391$) or including only low-risk ones ($g = 0.42$, $k = 55$). See Fig. S13 a raincloud plot and Fig. S14 for a descriptive specification curve plot focusing on including only low risk of bias studies.

Time Point. The methodological choice of including only effect sizes measured directly post-intervention or including only follow-up assessments (>24 weeks), resulted in statistically different effect size estimates. Post-intervention outcomes ($g = 0.46$, $k = 3091$) were larger than follow-up assessments ($g = 0.27$, $k = 547$). See Fig. S15 for a raincloud plot and Fig. S16 for a descriptive specification curve plot focusing only on follow-up meta-analyses.

Income level. Meta-analyses that exclusively included primary studies from low- and middle-income countries were associated with larger effect sizes ($g = 0.62$, $k = 58$) than meta-analyses where all studies were included, irrespective of income level ($g = 0.43$, $k = 3580$). See Fig. S17 a raincloud plot and Fig. S18 for a descriptive specification curve plot focusing on including only low risk of bias studies.

Meta-Analytical Method. Most meta-analytical methods yielded very similar effect size estimates, except PET-PEESE. The largest effect sizes were found when meta-analyses used p-uniform, $g = 0.52$, 95 % *CI* [0.26, 0.76] with $k = 359$ meta-analyses, while the smallest summary effect sizes were found with PET-PEESE, median $g = 0.23$, 95 % *CI* [−0.16, 0.62] with $k = 375$. All other meta-analytic estimators yielded results very similar to the average summary effect size of g = 0.43, 95 % *CI* [−0.16, 0.74]. See Fig. S19 for a raincloud plot and Fig. S20 for a descriptive specification curve highlighting meta-analyses that were analysed using the WAAP method.

Additionally, the investigation of the eight methods revealed no evidence of a Janus effect (see Table 2). Effect sizes were consistently positive at both the 90th and 10th percentiles, regardless of whether a specific estimator was included or excluded. However, the PET-PEESE estimators approached zero on their own, rendering their results ambiguous.

4. Discussion

4.1. Main findings

Our extensive multiverse meta-analysis, encompassing 3638 meta-analyses based on 125 RCTs and 263 effect sizes, provides evidence supporting the effectiveness of digital interventions for the treatment of depression. This finding is reinforced by positive effect sizes observed at

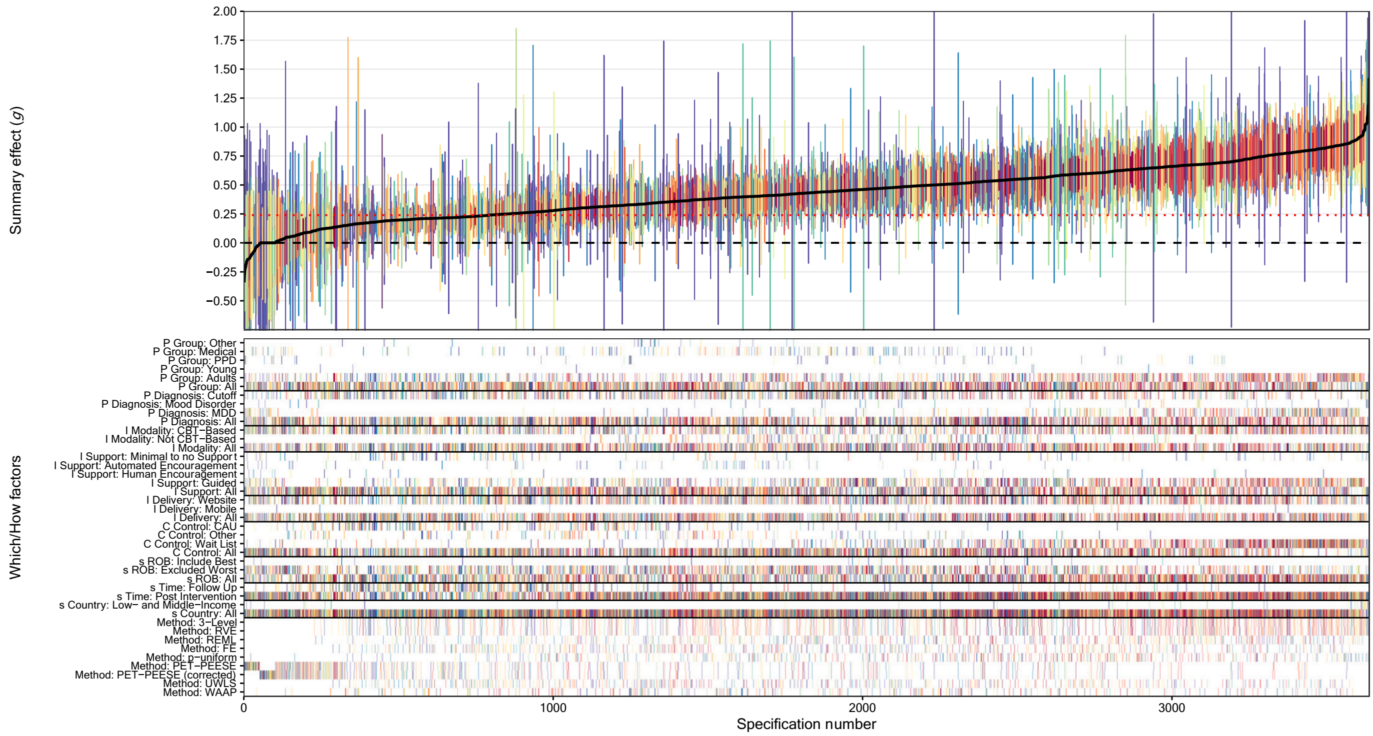

Fig. 2. Descriptive specification curve plot of all possible meta-analyses on digital interventions for depression disorders.
Note. The top panel shows the meta-analytic summary effects (expressed as Hedges' g) of all 3638 meta-analyses based on all possible analytic decisions with their 95 % confidence intervals. The summary effects are sorted by their magnitude, ranging from lower to higher effect size estimates from left to right. Each confidence interval is colour-coded, signifying the number of samples included in a specification (hot spectral colours indicate more included samples vs. cool spectral colours for less included samples). Connecting the different summary effects results in the solid black line, which is the specification curve. A horizontal dashed line of no effect is shown at $g = 0$ and a red dotted line is shown to indicate a small, yet clinically relevant effect size at $g = 0.24$. The overall pattern of the specification curve indicates that most meta-analyses produced summary effect size estimates larger than both horizontal lines. The vertical columns in the bottom panel represent factor combinations of all *Which* and *How* factors. These include different populations (target populations, types of depression assessments), interventions (formats, levels of support, treatment modalities), control conditions, as well as study features (such as strategies to deal with the risk of bias, time of assessment, and income level of country). The *How* factors represent eight meta-analytical estimators: 3-level meta-analytical models, RVE = robust variance estimation, REML = restricted maximum likelihood estimation, FE = fixed effects model, p-uniform*, PET-PEESE = Precision-Effect Test and Precision-Effect Estimate with Standard Errors, UWLS = Unrestricted Weighted Least Squares, and WAAP = Weighted Average of Adequately Powered. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

both the 10th (Hedges' $g = 0.16$) and 90th percentiles (Hedges' $g = 0.74$), as well as a median effect size of Hedges' $g = 0.43$. Furthermore, a substantial majority (97 %) of the effect sizes were >0, with 78 % reaching a clinically relevant effect size of Hedges' $g > 0.24$, indicating a robust and relevant effect across various methodological decisions. The overall analyses also revealed that most of the meta-analyses (82 %) showed a statistically significant difference in favour of digital interventions, while no meta-analysis significantly favoured the control condition.

As expected, the results showed variations based on several *Which* factors such as the level of support provided, the technology used, the populations targeted, and the control conditions employed. As an example, human-guided interventions, comparison with waitlist control groups, and mobile-based formats, were associated with larger effect sizes, whereas care as usual as a control condition, longer follow-up durations, including only low risk of bias studies, and focusing on populations with specific demographic characteristics such as medical groups or migrants were linked with smaller effect sizes. The investigation of the Vibrations of Effects further emphasises the consistency of these findings across different numbers of trials, reducing concerns about the robustness of the digital interventions' effectiveness. In summary, this comprehensive analysis underscores the positive impact of digital interventions for depression, while also showcasing the factors associated with their effectiveness as reported by all possible meta-analyses.

4.1.1. *Research in context*

To place our findings within the broader research landscape, previous meta-analyses reported variations in effect sizes for digital interventions for depression and have ranged from 0.22 to 1.01 across different meta-analyses (Chan et al., 2022; Cuijpers et al., 2019; Firth et al., 2017; Han and Kim, 2022; Josephine et al., 2017; Karyotaki et al., 2017; Köhnen et al., 2021; Pang et al., 2021; Serrano-Ripoll et al., 2022; Sierra et al., 2018; Xiong et al., 2023). Our results, with a median effect size of 0.43 and an interquartile range between 0.26 and 0.60, are positioned within this spectrum, demonstrating consistency with previous findings while also contributing to the ongoing dialogue regarding the effectiveness of digital interventions. The pronounced variation in effect sizes observed in previous research, referred to as the "vibration of effects," is critically addressed in our analysis. We scrutinised what this vibration entails, assessing how different methodological paths and choices might contribute to these variances in reported outcomes. Our findings indicate a general stability in the direction of effect sizes, with a notable absence of the Janus effect, thus providing a more consolidated and reliable understanding of the impact of digital interventions on depression.

Overall, we found results that are consistent with patterns reported in other reviews of digital interventions for depression, but our effects were a bit smaller. For instance, the most recent meta-analysis on digital interventions for depression by Moshe et al. (2021) reported modest effect sizes (Hedges' $g = 0.52$), larger effect sizes for human-guided

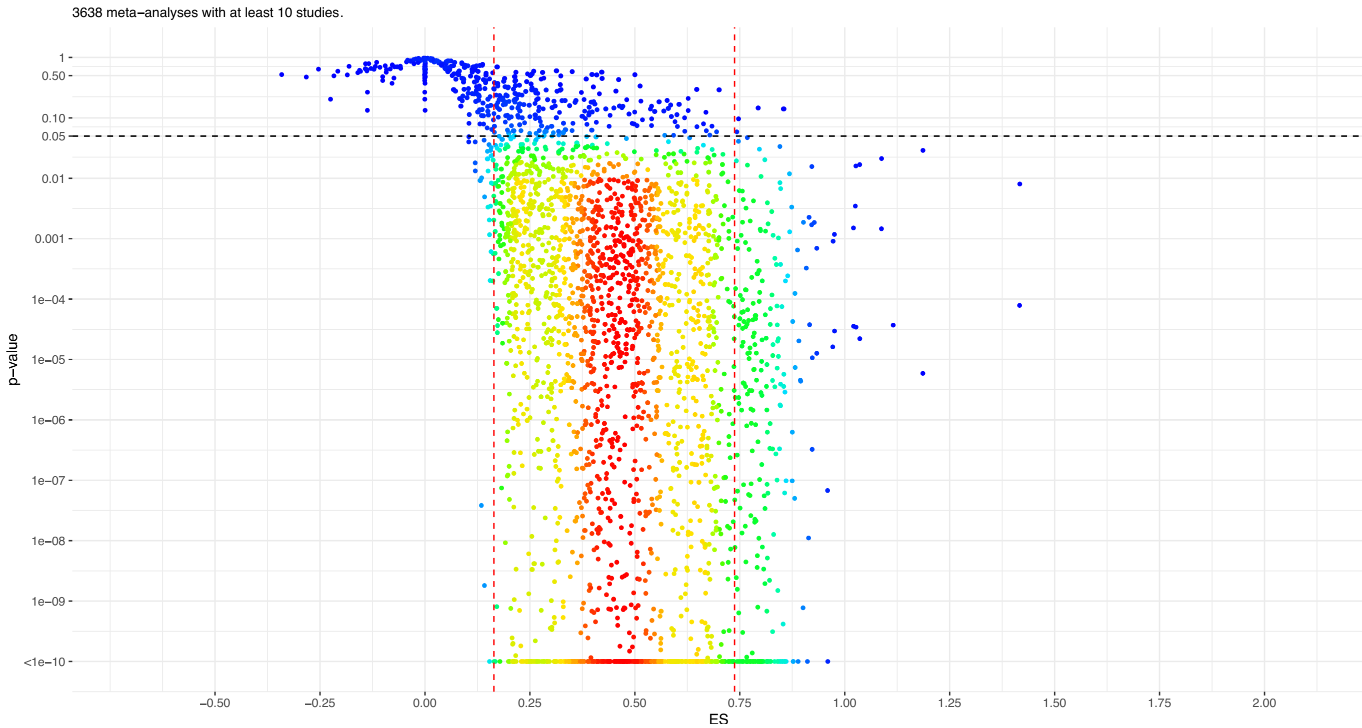

Fig. 3. Vibration of effects for the comparison of digital interventions with control conditions.
Note. The heat map shows the summary effect sizes (Hedges' *g*) of all 3638 meta-analyses and their respective *p*-values. A negative effect size suggests a higher effectiveness in reducing depression for the control group, and a positive effect size favours the digital interventions. Each point represents the meta-analyses. The colours represent the densities. The analysis is presented for meta-analyses containing at least 10 primary studies and in the supplement for meta-analyses including at least 2, 5, 25 or 50 primary studies. The vertical red dotted lines show the 10th and 90th percentiles and the horizontal black dotted line shows the *p*-value threshold at alpha = 0.05. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

interventions than in self-help interventions, larger effects when interventions were compared with waiting list control groups, and lower effects when compared with treatment as usual. Additionally, our findings align with those from a multiverse meta-analysis on psychotherapy effectiveness for depression, which showed that: a) interventions compared with waiting list control groups tend to show greater effectiveness; b) the inclusion of studies with a high risk of bias tends to inflate effect sizes; and c) adjustments for publication bias typically reduce reported effects (Plessen et al., 2023).

Our analysis has also enabled the performance of additional, previously unexplored meta-analyses, broadening the scope of our understanding of digital interventions in varying demographic groups and populations. For instance, we have identified patterns across different demographics, including that meta-analyses investigating adult populations tend to produce higher summary effects than those in medical groups, other groups, and younger populations—yet these smaller effects remain still clinically meaningful. This comprehensive approach allows us to confidently extend our conclusions to diverse groups, offering a more inclusive and representative interpretation of the available data.

Our analysis further supports the evidence of a tiered effect of support levels on intervention outcomes, with fully guided interventions demonstrating the most pronounced efficacy (median $g = 0.51$). This suggests that direct, human-delivered support related to treatment content enhances treatment effectiveness more than automated encouragement (median $g = 0.32$) or minimal to no support (median $g = 0.37$). Even the modest addition of human encouragement without therapeutic guidance improves outcomes (median $g = 0.39$), aligning with the accountability-support model that posits the motivational value of personalised interaction (Mohr et al., 2011). These insights underline the potential for optimising digital interventions by strategically incorporating human elements to augment user engagement and program adherence.

Additionally, the observed vibration of effects associated with diagnostic criteria revealed that patients with a clinical diagnosis of Major Depressive Disorder (MDD) displayed larger treatment effects (median $g = 0.61$) compared to those identified through self-report measures. This trend echoes findings from Karyotaki et al. (2021), who reported that individuals with more severe baseline symptoms—typically present in clinically diagnosed cases—have higher remission rates. The discernible gradient in effect sizes across different diagnostic categories supports the necessity of considering baseline symptom severity in the assessment of treatment efficacy. Notably, the inclusion of broader diagnostic terms like Mood Disorder yielded intermediate effect sizes (median $g = 0.55$), suggesting that specificity in diagnosis correlates with greater treatment gains.

Our findings further indicate greater benefits in meta-analyses including only studies conducted in low and middle-income countries, demonstrating a medium to large effect size ($g = 0.66$) in favour of the intervention group. Consistent with previous evidence (Karyotaki et al., 2023; Kim et al., 2023), these results suggest that digital interventions can effectively address global mental health challenges and potentially reduce the treatment gap in resource-limited settings. This highlights their potential value as an important clinical tool for alleviating depressive symptoms. However, due to limited statistical power, we were unable to definitively ascertain whether the observed differences in effect sizes between LMICs and HICs were attributable to a higher risk of bias, the increased likelihood of using inactive controls in LMICs, or other contributing factors. It is important to note that a recent meta-analysis by Karyotaki et al. (2023) faced a similar challenge. This study was unable to conduct a sensitivity analysis restricted to studies with a low risk of bias due to the absence of such studies.

The general consistency of effect sizes across most meta-analytical methods reinforces the robustness of our findings, except for the PET-

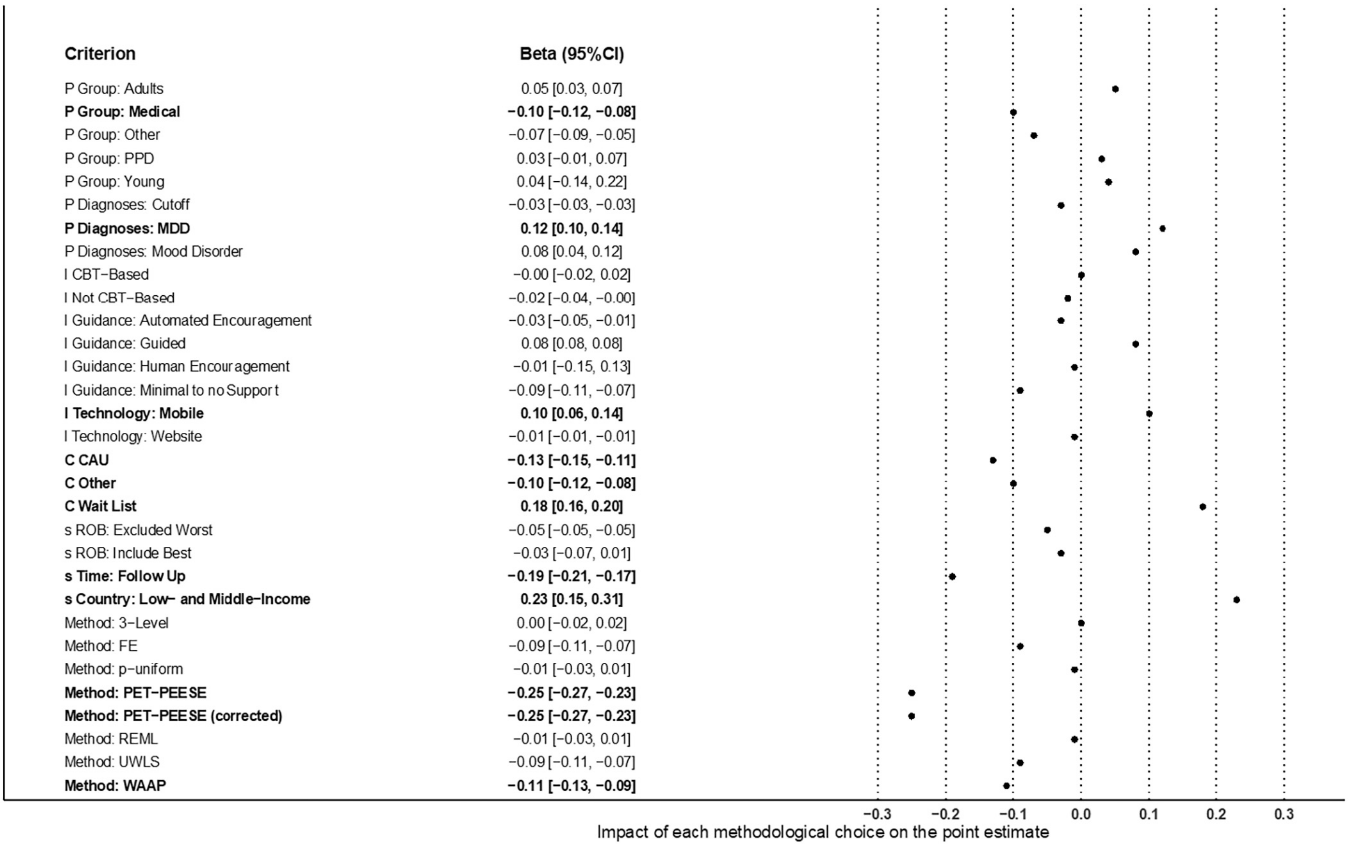

Fig. 4. Multiple median regression for each Which and How factor.
Note. Beta coefficients from median regressions for each *Which* and *How* Factor. This figure illustrates the impact of various methodological choices on the point estimate of effect size (Hedges' *g*) in a meta-analysis. The beta coefficients and their 95 % confidence intervals (CIs) are derived from median regressions for each specified methodological factor. The reference categories are comprehensive specifications that include all subgroups. The first letter of each criterion represents an element of the PICOs framework: P (Population), I (Intervention), C (Comparison), O (Outcome), s (study feature).. For example, "P Group" denotes different participant groups, "I Technology" denotes types of technology used in interventions, "C Wait List" denotes wait list comparison groups, and "s Time: Follow Up" denotes the study features related to follow-up duration.

PEESE estimator which produced notably lower estimates (median $g = 0.23$). This divergence suggests that PET-PEESE may offer a conservative lower threshold to the otherwise consistent effect size landscape.

Regarding the duration of treatment effectiveness, meta-analyses that included post-treatment assessments exhibited larger effect size estimates (median $g = 0.46$) compared to those reporting long-term outcomes, 24 weeks or more following the baseline assessment (median $g = 0.27$). Nevertheless, the long-term effectiveness remained above the clinically relevant cut-off score, indicating that digital interventions can prolong their effectiveness in the long term.

Lastly, our results suggest that CBT and non-CBT-based interventions have comparable effects, with effect size estimates showing minimal differences between the two modalities. This equivalence is reflected in the comparable effect sizes for non-CBT (median $g = 0.46$, $k = 272$) and CBT-based interventions (median $g = 0.43$, $k = 1402$), underscoring the potential for a range of therapeutic approaches to achieve similar outcomes in treating depressive disorders. Similar results were also found for interventions incorporating mobile-based (median $g = 0.47$, k = 58) or web-based technologies (median $g = 0.44$, $k = 1591$). However, the findings for treatment modalities and format of delivery should be interpreted with caution due to the disproportionate number of samples included in the meta-analyses.

The significance of our work lies not only in its contribution to the existing body of knowledge but also in its role in paving the way for future research and innovation in the field. By providing a robust and comprehensive analysis showcasing that overall, there is a robust modest effect for the effectiveness of digital interventions for depression, we establish a firm ground for further exploration and inquiry, highlighting areas that necessitate more attention and investigation. In particular, we acknowledge the necessity for future research to delve deeper into understanding the particular ingredients of effective digital interventions, specifically focusing on understanding the 'how' and 'for whom' aspects of these interventions. By doing so, we can further refine our understanding, enhance the effectiveness of digital interventions, and ensure that they are accessible and beneficial to all sections of the population.

4.2. Limitations

Our study has certain limitations. For one, the study was not preregistered. This decision was made due to the inherent type of multiverse meta-analyses in exploring all possible statistical decisions (how factors) and all possible combinations of *Which factors* as per our inclusion criteria. Second, the study quality of the original sample of 126 studies, rated with the risk of bias assessment tool (ROB2), was generally poor. Of all primary studies, 30 % were rated as high risk and only 13 % as low risk. Therefore, only a few meta-analyses could exclusively include low risk of bias studies. However, in those rare meta-analyses, only slightly lower effect size estimates were found than in meta-analyses including studies of poorer quality, indicating that the differences in study quality did not bias the meta-analytical findings.

Third, one common limitation of one of our methods to correct for small study effects, PET-PEESE, is its tendency to over-correct for biases, as highlighted by Carter et al. (2019). In our study, this issue is highly

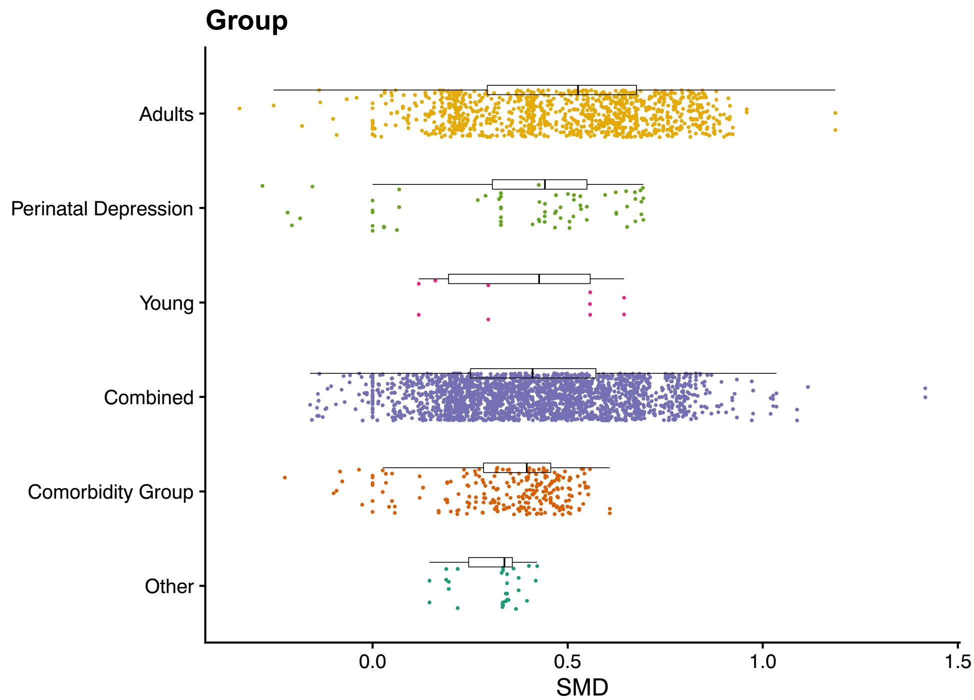

Fig. 5. Raincloud plot of all meta-analyses on digital interventions for depression, grouped by different populations.
Note. For each of the meta-analyses on different target populations, this raincloud plot shows the distribution with a boxplot (the cloud), and each individual meta-analysis (the rain).

likely: while many observed effect sizes are positive (93.92 %), the corrected summary effect sizes of meta-analyses can be substantially negative, even in meta-analyses where only positive effect sizes (20 meta-analyses analysed with PET-PEESE) or only one negative effect size (18 such meta-analyses) is present. To address this, we followed the recommendation to adjust these estimates to zero as these meta-analyses indicate that there is no significant effect. The results were overall very similar, indicating that this did not substantially distort our analyses.

Fourth, similar to conventional meta-analyses, one of the inherent challenges of conducting multiverse meta-analyses is the myriad of analytical decisions that must be made, each with the capability to impact the final outcomes. Although multiverse meta-analyses succeed in providing an extensive bird's eye view of a research field, much more so than conventional meta-analyses, the selection and specifications of *Which* and *How* factors are still subject to the individual researcher's decision. For example, had we opted for different methodologies, the results might have displayed distinct patterns, particularly in areas such as the descriptive specification curve and the VoE plot. A prime example of this variability can be observed when considering the use of diverse bias assessment tools or altering the length of follow-up durations, both of which can steer results in different directions. Multiverse meta-analyses, by design, seek to encapsulate a range of analytical scenarios to provide a holistic view of the research. For this reason, we aimed to include as many different methodological options as possible. Yet, this multiplicity can introduce a degree of variability that might cloud the precision and clarity of the findings. A potential avenue to mitigate this limitation and ensure more consistent results in future multiverse meta-analyses would be to integrate both systematic and umbrella reviews into the analytical framework (El Bahri et al., 2022).

Lastly, we performed a series of regressions to understand the sources of VoE. Although such an analysis must be interpreted cautiously because it is susceptible to confounders, it may help to understand the discrepancies found in multiple overlapping meta-analyses in a given field (El Bahri et al., 2022). Moreover, the VoE framework enables the a-posteriori inspection of some combinations of interest. Importantly, our quantification of the analytical decisions possibly associated with VoE is exploratory and must be interpreted with caution. A major limitation of our approach lies in the fact that we performed a regression on a wide range of meta-analyses, some of which are redundant (and are combinations of the same set of studies). Although point estimates are not expected to be biased, nonindependence between those meta-analyses prevents any easy computation of confidence intervals, which we did not report for this reason. Solutions to derive such confidence intervals (e.g., by weighting the results of the different studies or by analysing only unique combinations of studies) require more development before they are implemented and adopted widely (El Bahri et al., 2022).

4.3. *Strengths and future directions*

By using this broad array of models, we ensured that the findings are robust across different assumptions and addressed methodological issues that simple random effects models cannot handle. This enhances the rigor of our analysis and provides a more comprehensive understanding of the impact of various biases and dependencies in the data, contributing new insights to the evidence base.

Based on the identified decisions associated with systematically higher and lower effect sizes we were able to select distinct plausible combinations of different criteria, that would produce contradictory meta-analyses. For example, a very large summary effect can be obtained when conducting a meta-analysis focused on adults with MDD, comparing human guided interventions with wait list control groups, while keeping high risk of bias studies included (Hedges' $g = 0.91$, 95 % *CI* [0.67, 1.14], $k = 18$). On the other end of the spectrum, non-significant meta-analyses can be obtained when we conduct a meta-analysis focusing on minimal to no support guidance analysed with PET-PEESE (Hedges' $g = -0.03, 0.03, 0.13$, 95 % *CI* [−0.53, 0.46, −0.32, 0.37, −0.04, 0.29], $k = 10, 13, 17$).

Psychotherapy research has extensively explored the potential influences of many of these factors on the resulting effect sizes in meta-analyses, such as in different population, type of intervention, different control groups (Michopoulos et al., 2021), quality of trials (Cuijpers et al., 2010), publication bias (Driessen et al., 2015). Our

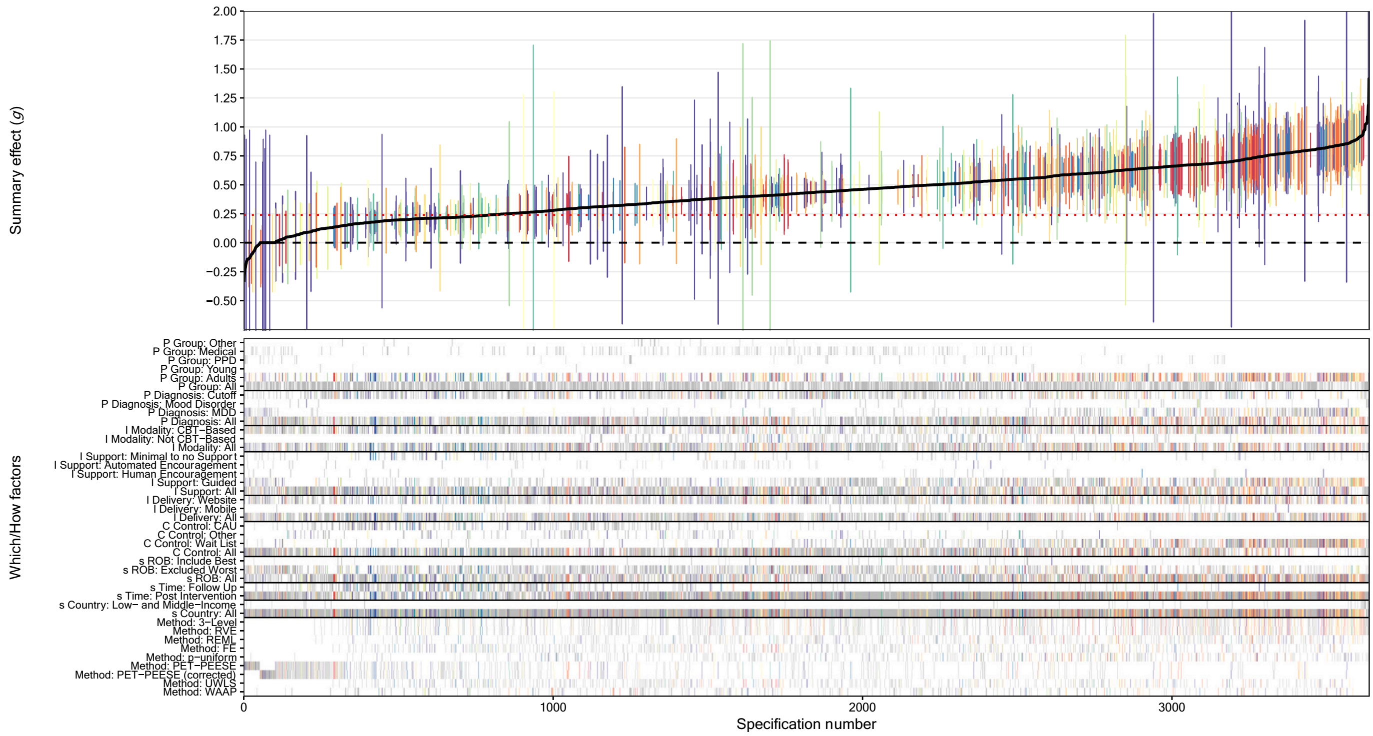

Fig. 6. Descriptive specification curve highlighting meta-analyses in adults.
Note. The top panel shows the meta-analytic summary effects (expressed as Hedges' *g*) of all 3638 meta-analyses based on all possible analytic decisions with their 95 % confidence intervals. The summary effects are sorted by their magnitude, ranging from lower to higher effect size estimates from left to right. Each confidence interval is colour-coded for meta-analyses focussed on the general adult population, signifying the number of samples included in a specification (hot spectral colours indicate more included samples vs. cool spectral colours for less included samples, while grey corresponds to other target populations). Connecting the different summary effects results in the solid black line, which is the specification curve. A horizontal dashed line of no effect is shown at $g = 0$ and a red dotted line is shown to indicate a small, yet clinically relevant effect size at $g = 0.24$. The overall pattern of the specification curve indicates that most meta-analyses produced summary effect size estimates larger than both horizontal lines. The vertical columns in the bottom panel represent factor combinations of all *Which* and *How* factors. These include different populations (target populations, types of depression assessments), interventions (formats, levels of support, treatment modalities), control conditions, as well as study features (such as strategies to deal with the risk of bias, time of assessment, and income level of country). The *How* factors represent eight meta-analytical estimators: 3-level meta-analytical models, RVE = robust variance estimation, REML = restricted maximum likelihood estimation, FE = fixed effects model, *p*-uniform*, PET-PEESE = Precision-Effect Test and Precision-Effect Estimate with Standard Errors, UWLS = Unrestricted Weighted Least Squares, and WAAP = Weighted Average of Adequately Powered. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

multiverse meta-analysis extended this exploration. Instead of examining these factors in isolation, it encompasses all analyses across these factors simultaneously, representing a vast range of plausible results. Such a comprehensive approach aids in understanding how different analytical choices, when combined, interact and sometimes lead to diverging conclusions.

Addressing the sometimes-conflicting results generated by the recent surge in meta-analysis production, this study's application of the multiverse meta-analysis approach is particularly helpful. The evident vibrations of effects from similar research questions emphasise the critical need for innovative methodologies capable of tackling the reproducibility challenges inherent in meta-analyses. Identifying the factors contributing to these vibrations has been a key outcome of our study. For instance, we found that when investigating self-guided or comparing treatment as usual control groups, the intervention was deemed less effective, aligning with findings from previous studies (Cuijpers et al., 2019; Pang et al., 2021). In contrast, guided interventions have consistently shown larger effect sizes over self-guided ones, as supported by multiple studies (Mamukashvili-Delau et al., 2022; Wells et al., 2018; Wright et al., 2019), while other studies found guided and self-guided interventions to be equivalent (Ahern et al., 2018; Sztein et al., 2018). Moreover, our analysis underscores the impact of follow-up timepoints and different approaches to handling the risk of bias, further contributing to the complexity and inconsistency of findings across different meta-analyses. This nuanced understanding highlights the necessity for careful consideration and transparency in methodological choices to enhance the reliability and validity of meta-analytic findings in the field of digital interventions for depression.

Consequently, clinicians, researchers, and policymakers should exercise caution and employ a rigorous, critical lens when interpreting and applying these findings (De Vrieze, 2018). This vigilant approach ensures that decisions are rooted in a thorough understanding of the underlying data and its potential variations.

The effectiveness of these interventions is especially meaningful in light of the high prevalence of depression, and the associated burdens—both personal and economic (Johnston et al., 2019; Santomauro et al., 2021; World Health Organization, 2023).The increased accessibility and immediacy of digital interventions present an important avenue for support, especially given the existing shortages in available therapy programs (Cuijpers and Riper, 2015; Patel et al., 2020; Rodriguez-Villa et al., 2020). This is particularly relevant for individuals who find themselves on lengthy waiting lists, as these digital tools can offer much-needed assistance in the interim. While preliminary evidence suggests attitudes towards digital interventions have become more positive in recent years (Kim et al., 2023), there remains a need to further develop and integrate effective treatment strategies that address individual needs, motivation and perceptions towards treatment, to enhance treatment acceptability (Chan and Honey, 2022; Zhong et al., 2023). As we navigate the complexities of mental health care, the incorporation of digital interventions emerges as a vital component,

Table 2
Janus effect and effect size information for meta-analyses with at least 10 primary studies.

Method	Type	10th	90th	Median	Mean	*k*
REML	Only Method	0.26	0.77	0.50	0.51	368
REML	Excluded	0.15	0.73	0.42	0.43	3270
FE	Only Method	0.22	0.68	0.42	0.44	370
FE	Excluded	0.15	0.74	0.43	0.44	3268
P-UNIFORM	Only Method	0.28	0.75	0.52	0.52	359
P-UNIFORM	Excluded	0.15	0.74	0.42	0.43	3279
PET-PEESE	Only Method	0.00	0.55	0.20	0.25	375
PET-PEESE	Excluded	0.20	0.75	0.45	0.46	3263
UWLS	Only Method	0.22	0.68	0.42	0.44	375
UWLS	Excluded	0.15	0.74	0.43	0.44	3263
WAAP	Only Method	0.21	0.69	0.41	0.43	324
WAAP	Excluded	0.16	0.74	0.43	0.44	3314
3-LEVEL	Only Method	0.22	0.80	0.50	0.51	546
3-LEVEL	Excluded	0.14	0.72	0.41	0.42	3092
RVE	Only Method	0.22	0.80	0.50	0.51	546
RVE	Excluded	0.14	0.72	0.41	0.42	3092
PET-PEESE (CORRECTED)	Only Method	0.00	0.55	0.20	0.25	375
PET-PEESE (CORRECTED)	Excluded	0.20	0.75	0.45	0.46	3263
ALL	All	0.16	0.74	0.43	0.44	3638

Notes. REML: Restricted Maximum Likelihood; FE: Fixed/Equal Effect Model; P-UNIFORM: P-Uniform Method; PET-PEESE: Precision Effect Test and Precision Effect Estimate with Standard Error; UWLS: Unweighted Least Squares; WAAP: Weighted Average Absolute Prediction; 3-LEVEL: Three-level Meta-analysis; RVE: Robust Variance Estimation. Type: *Only Method* contains all meta-analyses based on that specific method while *Excluded* contains all other methods except that specific method.

contributing to a more accessible system, ready to meet the diverse needs of its users.

In summary, our multiverse meta-analysis not only reaffirms the effectiveness of digital interventions in treating depression but also enriches the discourse by providing a more nuanced, comprehensive, and inclusive understanding of their impact. Our work stands as a testament to the potential of digital interventions, serving as an important avenue for support and advancement in the field of mental health, while also setting the stage for future research endeavours aimed at optimising and personalising digital mental health care.

Role of the funding source

No funding was provided for this study.

CRediT authorship contribution statement

Constantin Yves Plessen: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Olga Maria Panagiotopoulou:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Lingyao Tong:** Data curation, Investigation, Writing – original draft, Writing – review & editing. **Pim Cuijpers:** Writing – review & editing, Supervision, Conceptualization. **Eirini Karyotaki:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2024.10.018.

References

Ahern, E., Kinsella, S., Semkovska, M., 2018. Clinical efficacy and economic evaluation of online cognitive behavioral therapy for major depressive disorder: a systematic review and meta-analysis. Expert Rev. Pharmacoecon. Outcomes Res. 18 (1), 25–41. https://doi.org/10.1080/14737167.2018.1407245.

Allen, M., Poggiali, D., Whitaker, K., Marshall, T.R., van Langen, J., Kievit, R.A., 2021. Raincloud plots: a multi-platform tool for robust data visualization. Wellcome Open Res. 4, 63. https://doi.org/10.12688/wellcomeopenres.15191.2.

American Psychological Association, 2019. Clinical practice guideline for the treatment of depression across three age cohorts. https://www.apa.org/depression-guideline.

Carbonell, Á., Navarro-Pérez, J., Mestre, M., 2020. Challenges and barriers in mental healthcare systems and their impact on the family: a systematic integrative review. Health Soc. Care Commun. 28 (5), 1366–1379. https://doi.org/10.1111/hsc.12968.

Carter, E.C., Schönbrodt, F.D., Gervais, W.M., Hilgard, J., 2019. Correcting for bias in psychology: a comparison of meta-analytic methods. Adv. Methods Pract. Psychol. Sci. 2 (2), 115–144. https://doi.org/10.1177/2515245919847196.

Chan, A.H.Y., Honey, M.L.L., 2022. User perceptions of mobile digital apps for mental health: Acceptability and usability - An integrative review. J. Psych. Mental Health Nurs. 29 (1), 147–168. https://doi.org/10.1111/jpm.12744.

Chan, M., Jiang, Y., Lee, C.Y.C., Ramachandran, H.J., Teo, J.Y.C., Seah, C.W.A., Lin, Y., Wang, W., 2022. Effectiveness of eHealth-based cognitive behavioural therapy on depression: a systematic review and meta-analysis. J. Clin. Nurs. 31 (21–22), 3021–3031. https://doi.org/10.1111/jocn.16212.

Cuijpers, P., Riper, H., 2015. [Internet interventions for depressive disorders: an overview]. Intervenciones basadas en internet para los trastornos depresivos: Una visión general. Rev. Psicopatol. Psicol. Clin. 19 (3), 209. https://doi.org/10.5944/rppc.vol.19.num.3.2014.13902.

Cuijpers, P., Van Straten, A., Warmerdam, L., Andersson, G., 2008. Psychological treatment of depression: a meta-analytic database of randomized studies. BMC Psychiatry 8 (1), 36. https://doi.org/10.1186/1471-244X-8-36.

Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S.D., Andersson, G., 2010. The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. Psychol. Med. 40 (2), 211–223. https://doi.org/10.1017/S0033291709006114.

Cuijpers, P., Turner, E.H., Koole, S.L., Van Dijke, A., Smit, F., 2014. What is the threshold for a clinically relevant effect? The case of major depressive disorders: theoretical review: threshold for a clinically relevant effect. Depress. Anxiety 31 (5), 374–378. https://doi.org/10.1002/da.22249.

Cuijpers, P., Noma, H., Karyotaki, E., Cipriani, A., Furukawa, T.A., 2019. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. JAMA Psychiatry 76 (7), 700. https://doi.org/10.1001/jamapsychiatry.2019.0268.

Cuijpers, P., Karyotaki, E., Ciharova, M., Quero, S., Pineda, B., Munoz, R., Rosenström, T., 2022. A Meta-analytic Database of Randomised Trials on Psychotherapies for Depression. https://doi.org/10.17605/OSF.IO/825C6.

De Vrieze, J., 2018. The metawars. Science 361 (6408), 1184–1188. https://doi.org/10.1126/science.361.6408.1184.

Driessen, E., Hollon, S.D., Bockting, C.L.H., Cuijpers, P., Turner, E.H., 2015. Does publication bias inflate the apparent efficacy of psychological treatment for major depressive disorder? A systematic review and meta-analysis of US National Institutes of Health-funded trials. PLoS One 10 (9), e0137864. https://doi.org/10.1371/journal.pone.0137864.

Ebert, D.D., Cuijpers, P., Muñoz, R.F., Baumeister, H., 2017. Prevention of mental health disorders using internet- and mobile-based interventions: a narrative review and recommendations for future research. Front. Psychiatry 8, 116. https://doi.org/10.3389/fpsyt.2017.00116.

El Bahri, M., Wang, X., Biaggi, T., Falissard, B., Naudet, F., Barry, C., 2022. A multiverse analysis of meta-analyses assessing acupuncture efficacy for smoking cessation evidenced vibration of effects. J. Clin. Epidemiol. 152, 140–150. https://doi.org/10.1016/j.jclinepi.2022.09.001.

Evans-Lacko, S., Knapp, M., 2016. Global patterns of workplace productivity for people with depression: absenteeism and presenteeism costs across eight diverse countries. Soc. Psychiatry Psychiatr. Epidemiol. 51 (11), 1525–1537. https://doi.org/10.1007/s00127-016-1278-4.

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., Sarris, J., 2017. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 16 (3), 287–298. https://doi.org/10.1002/wps.20472.

Fried, E.I., Nesse, R.M., 2014. The impact of individual depressive symptoms on impairment of psychosocial functioning. PLoS One 9 (2), e90311. https://doi.org/10.1371/journal.pone.0090311.

Furukawa, T.A., Suganuma, A., Ostinelli, E.G., Andersson, G., Beevers, C.G., Shumake, J., Berger, T., Boele, F.W., Buntrock, C., Carlbring, P., Choi, I., Christensen, H., Mackinnon, A., Dahne, J., Huibers, M.J.H., Ebert, D.D., Farrer, L., Forand, N.R., Strunk, D.R., et al., 2021. Dismantling, optimising, and personalising internet cognitive behavioural therapy for depression: a systematic review and component network meta-analysis using individual participant data. Lancet Psychiatry 8 (6), 500–511. https://doi.org/10.1016/S2215-0366(21)00077-8.

Han, A., Kim, T.H., 2022. Effects of internet-delivered behavioral activation on individuals with depressive symptoms: a systematic review and meta-analysis. J. Psychiatr. Res. 152, 104–118. https://doi.org/10.1016/j.jpsychires.2022.05.031.

Harrer, M., Cuijpers, P., Furukawa, T., Ebert, D.D., 2021. Doing Meta-analysis With R: A Hands-on Guide, 1st ed. Chapman & Hall/CRC Press.

Harrer, M., Kuper, P., Sprenger, A., Cuijpers, P., 2022. *metapsyTools: Several R Helper Functions For the "Metapsy" Database* (0.3.3-5) [Computer Software]. Zenodo. https://doi.org/10.5281/ZENODO.6566632.

Hedman-Lagerlöf, E., Carlbring, P., Svärdman, F., Riper, H., Cuijpers, P., Andersson, G., 2023. Therapist-supported Internet-based cognitive behaviour therapy yields similar effects as face-to-face therapy for psychiatric and somatic disorders: An updated systematic review and meta-analysis. World Psych. 22 (2), 305–314.

Higinbotham, M.K., Emmert-Aronson, B., Bunge, E.L., 2020. A meta-analysis of the effectiveness of behavioral intervention technologies and face-to-face cognitive behavioral therapy for youth with depression. J. Technol. Behav. Sci. 5 (4), 324–335. https://doi.org/10.1007/s41347-020-00139-6.

Jain, S., Gupta, S., Li, V.W., Suthoff, E., Arnaud, A., 2022. Humanistic and economic burden associated with depression in the United States: a cross-sectional survey analysis. BMC Psychiatry 22 (1), 542. https://doi.org/10.1186/s12888-022-04165-x.

Johnston, D.A., Harvey, S.B., Glozier, N., Calvo, R.A., Christensen, H., Deady, M., 2019. The relationship between depression symptoms, absenteeism and presenteeism. J. Affect. Disord. 256, 536–540. https://doi.org/10.1016/j.jad.2019.06.041.

Josephine, K., Josefine, L., Philipp, D., David, E., Harald, B., 2017. Internet- and mobile-based depression interventions for people with diagnosed depression: a systematic review and meta-analysis. J. Affect. Disord. 223, 28–40. https://doi.org/10.1016/j.jad.2017.07.021.

Karyotaki, E., Riper, H., Twisk, J., Hoogendoorn, A., Kleiboer, A., Mira, A., Mackinnon, A., Meyer, B., Botella, C., Littlewood, E., Andersson, G., Christensen, H., Klein, J.P., Schröder, J., Bretón-López, J., Scheider, J., Griffiths, K., Farrer, L., Huibers, M.J.H., et al., 2017. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 74 (4), 351. https://doi.org/10.1001/jamapsychiatry.2017.0044.

Karyotaki, E., Efthimiou, O., Miguel, C., Bermpohl, F.M.G., Furukawa, T.A., Cuijpers, P., Riper, H., Patel, V., Mira, A., Gemmil, A.W., Yeung, A.S., Lange, A., Williams, A.D., Mackinnon, A., Geraedts, A., Van Straten, A., Meyer, B., Björkelund, C., et al., Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, 2021. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 78 (4), 361. https://doi.org/10.1001/jamapsychiatry.2020.4364.

Karyotaki, E., Miguel, C., Panagiotopoulou, O.M., Harrer, M., Seward, N., Sijbrandij, M., Araya, R., Patel, V., Cuijpers, P., 2023. Digital interventions for common mental disorders in low- and middle-income countries: A systematic review and meta-analysis. Global Mental Health (Cambridge, England) 10, e68. https://doi.org/10.1017/gmh.2023.50.

Karyotaki, E., Panagiotopoulou, O.-M., Tong, L., Amarnath, A., Cuijpers, P., 2024. Digital mental health interventions for common mental health disorders (The SYNTHESIS Project): An Individual Patient Data Repository. https://doi.org/10.17605/OSF.IO/P5EMR.

Kheirkhah, F., Faramarzi, M., Shafierizi, S., Chehrazi, M., Basirat, Z., 2023. Preliminary examination of acceptability, feasibility, and effectiveness of internet-based cognitive behavioral therapy for treatment of depression and anxiety in infertile women. Heliyon 9 (5), e15760. https://doi.org/10.1016/j.heliyon.2023.e15760.

Kim, J., Aryee, L.M.D., Bang, H., Prajogo, S., Choi, Y.K., Hoch, J.S., Prado, E.L., 2023. Effectiveness of digital mental health tools to reduce depressive and anxiety symptoms in low- and middle-income countries: Systematic review and meta-analysis. JMIR Mental Health 10, e43066. https://doi.org/10.2196/43066.

Köhnen, M., Kriston, L., Härter, M., Baumeister, H., Liebherz, S., 2021. Effectiveness and acceptance of technology-based psychological interventions for the acute treatment of unipolar depression: systematic review and meta-analysis. J. Med. Internet Res. 23 (6). https://doi.org/10.2196/24584.

Laursen, T.M., Musliner, K.L., Benros, M.E., Vestergaard, M., Munk-Olsen, T., 2016. Mortality and life expectancy in persons with severe unipolar depression. J. Affect. Disord. 193, 203–207. https://doi.org/10.1016/j.jad.2015.12.067.

Lindegaard, T., Berg, M., Andersson, G., 2020. Efficacy of internet-delivered psychodynamic therapy: systematic review and meta-analysis. Psychodyn. Psychiatry 48 (4), 437–454. https://doi.org/10.1521/pdps.2020.48.4.437.

Mamukashvili-Delau, M., Koburger, N., Dietrich, S., Rummel-Kluge, C., 2022. Efficacy of computer- and/or internet-based cognitive-behavioral guided self-management for depression in adults: a systematic review and meta-analysis of randomized controlled trials. BMC Psychiatry 22 (1), 730. https://doi.org/10.1186/s12888-022-04325-z.

Michopoulos, I., Furukawa, T.A., Noma, H., Kishimoto, S., Onishi, A., Ostinelli, E.G., Ciharova, M., Miguel, C., Karyotaki, E., Cuijpers, P., 2021. Different control conditions can produce different effect estimates in psychotherapy trials for depression. J. Clin. Epidemiol. 132, 59–70. https://doi.org/10.1016/j.jclinepi.2020.12.012.

Mohr, D.C., Hart, S.L., Howard, I., Julian, L., Vella, L., Catledge, C., Feldman, M.D., 2006. Barriers to psychotherapy among depressed and nondepressed primary care patients. Ann. Behav. Med. 32 (3), 254–258. https://doi.org/10.1207/s15324796abm3203_12.

Mohr, D.C., Cuijpers, P., Lehman, K., 2011. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J. Med. Internet Res. 13 (1), e30. https://doi.org/10.2196/jmir.1602.

Moitra, M., Santomauro, D., Collins, P.Y., Vos, T., Whiteford, H., Saxena, S., Ferrari, A.J., 2022. The global gap in treatment coverage for major depressive disorder in 84 countries from 2000–2019: a systematic review and Bayesian meta-regression analysis. PLoS Med. 19 (2), e1003901. https://doi.org/10.1371/journal.pmed.1003901.

Moshe, I., Terhorst, Y., Philippi, P., Domhardt, M., Cuijpers, P., Cristea, I., Pulkki-Råback, L., Baumeister, H., Sander, L.B., 2021. Digital interventions for the treatment of depression: a meta-analytic review. Psychol. Bull. 147 (8), 749–786. https://doi.org/10.1037/bul0000334.

Pang, Y., Zhang, X., Gao, R., Xu, L., Shen, M., Shi, H., Li, Y., Li, F., 2021. Efficacy of web-based self-management interventions for depressive symptoms: a meta-analysis of randomized controlled trials. BMC Psychiatry 21 (1), 398. https://doi.org/10.1186/s12888-021-03396-8.

Patel, C.J., Burford, B., Ioannidis, J.P.A., 2015. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. J. Clin. Epidemiol. 68 (9), 1046–1058. https://doi.org/10.1016/j.jclinepi.2015.05.029.

Patel, S., Akhtar, A., Malins, S., Wright, N., Rowley, E., Young, E., Sampson, S., Morriss, R., 2020. The acceptability and usability of digital health interventions for adults with depression, anxiety, and somatoform disorders: qualitative systematic review and meta-synthesis. J. Med. Internet Res. 22 (7), e16228. https://doi.org/10.2196/16228.

Plessen, C.Y., Karyotaki, E., Miguel, C., Ciharova, M., Cuijpers, P., 2023. Exploring the efficacy of psychotherapies for depression: a multiverse meta-analysis. BMJ Mental Health 26 (1), e300626. https://doi.org/10.1136/bmjment-2022-300626.

Reins, J.A., Buntrock, C., Zimmermann, J., Grund, S., Harrer, M., Lehr, D., Baumeister, H., Weisel, K., Domhardt, M., Imamura, K., Kawakami, N., Spek, V., Nobis, S., Snoek, F., Cuijpers, P., Klein, J.P., Moritz, S., Ebert, D.D., 2021. Efficacy and moderators of internet-based interventions in adults with subthreshold depression: an individual participant data meta-analysis of randomized controlled trials. Psychother. Psychosom. 90 (2), 94–106. https://doi.org/10.1159/000507819.

Rodriguez-Villa, E., Naslund, J., Keshavan, M., Patel, V., Torous, J., 2020. Making mental health more accessible in light of COVID-19: scalable digital health with digital navigators in low and middle-income countries. Asian J. Psychiatr. 54, 102433. https://doi.org/10.1016/j.ajp.2020.102433.

Roman, M., Constantin, T., Bostan, C.M., 2020. The efficiency of online cognitive-behavioral therapy for postpartum depressive symptomatology: a systematic review and meta-analysis. Women Health 60 (1), 99–112. https://doi.org/10.1080/03630242.2019.1610824.

Santomauro, D.F., Mantilla Herrera, A.M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D.M., Abbafati, C., Adolph, C., Amlag, J.O., Aravkin, A.Y., Bang-Jensen, B.L., Bertolacci, G.J., Bloom, S.S., Castellano, R., Castro, E., Chakrabarti, S., Chattopadhyay, J., Cogen, R.M., Collins, J.K., et al., 2021. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 398 (10312), 1700–1712. https://doi.org/10.1016/S0140-6736(21)02143-7.

Schaffler, Y., Probst, T., Jesser, A., Humer, E., Pieh, C., Stippl, P., Haid, B., Schigl, B., 2022. Perceived barriers and facilitators to psychotherapy utilisation and how they relate to patient's psychotherapeutic goals. Healthcare 10 (11), 2228. https://doi.org/10.3390/healthcare10112228.

Serrano-Ripoll, M.J., Zamanillo-Campos, R., Fiol-DeRoque, M.A., Castro, A., Ricci-Cabello, I., 2022. Impact of smartphone app–based psychological interventions for reducing depressive symptoms in people with depression: systematic literature review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 10 (1), e29621. https://doi.org/10.2196/29621.

Sierra, M.A., Ruiz, F.J., Fundación Universitaria Konrad Lorenz, Flórez, C.L., Fundación Universitaria Konrad Lorenz, 2018. A systematic review and meta-analysis of third-wave online interventions for depression. Rev. Latinoam. Psicol. 50 (2). https://doi.org/10.14349/rlp.2018.v50.n2.6.

Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psych. Sci. 22 (11), 1359–1366.

Simonsohn, U., Simmons, J.P., Nelson, L.D., 2020. Specification curve analysis. Nat. Human Behav. 4 (11), 1208–1214.

Stanley, T.D., 2017. Limitations of PET-PEESE and other meta-analysis methods. Soc. Psychol. Personal. Sci. 8 (5), 581–591. https://doi.org/10.1177/1948550617693062.

Stanley, T.D., Doucouliagos, H., 2014. Meta-regression approximations to reduce publication selection bias. Res. Synth. Methods 5 (1), 60–78.

Stanley, T.D., Doucouliagos, H., Ioannidis, J.P.A., 2022. Beyond random effects: when small-study findings are more heterogeneous. Adv. Methods Pract. Psychol. Sci. 5 (4), 251524592211204. https://doi.org/10.1177/25152459221120427.

Steegen, S., Tuerlinckx, F., Gelman, A., Vanpaemel, W., 2016. Increasing transparency through a multiverse analysis. Persp. Psych. Sci. 11 (5), 702–712.
Sterne, J.A.C., Savović, J., Page, M.J., Elbers, R.G., Blencowe, N.S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M.S., Eldridge, S.M., Emberson, J.R., Hernán, M.A., Hopewell, S., Hróbjartsson, A., Junqueira, D.R., Jüni, P., Kirkham, J.J., Lasserson, T., Li, T., et al., 2019. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ l4898. https://doi.org/10.1136/bmj.l4898.
Stewart, W.F., 2003. Cost of lost productive work time among US workers with depression. JAMA 289 (23), 3135. https://doi.org/10.1001/jama.289.23.3135.
Sztein, D.M., Koransky, C.E., Fegan, L., Himelhoch, S., 2018. Efficacy of cognitive behavioural therapy delivered over the internet for depressive symptoms: a systematic review and meta-analysis. J. Telemed. Telecare 24 (8), 527–539. https://doi.org/10.1177/1357633X17717402.
Van Aert, R.C.M., Van Assen, M.A.L.M., 2018. *Correcting for Publication Bias in a Meta-analysis With the P-uniform* Method* [Preprint]. MetaArXiv. https://doi.org/10.31222/osf.io/zqjr9.
Viechtbauer, W., 2010. Conducting meta-analyses in *R* with the **metafor** package. J. Stat. Softw. 36 (3). https://doi.org/10.18637/jss.v036.i03.
Voracek, M., Kossmeier, M., Tran, U.S., 2019. Which data to meta-analyze, and how?: a specification-curve and multiverse-analysis approach to meta-analysis. Z. Psychol. 227 (1), 64–82. https://doi.org/10.1027/2151-2604/a000357.
Wells, M.J., Owen, J.J., McCray, L.W., Bishop, L.B., Eells, T.D., Brown, G.K., Richards, D., Thase, M.E., Wright, J.H., 2018. Computer-assisted cognitive-behavior therapy for depression in primary care: systematic review and meta-analysis. Prim. Care Companion CNS Disord. 20 (2). https://doi.org/10.4088/PCC.17r02196.
Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij, J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., et al., 2016. The FAIR guiding principles for scientific data management and stewardship. Sci. Data 3 (1), 160018. https://doi.org/10.1038/sdata.2016.18.
World Health Organization, 2023. Depressive disorder (depression). https://www.who.int/news-room/fact-sheets/detail/depression.
Wright, J.H., Owen, J.J., Richards, D., Eells, T.D., Richardson, T., Brown, G.K., Barrett, M., Rasku, M.A., Polser, G., Thase, M.E., 2019. Computer-assisted cognitive-behavior therapy for depression: a systematic review and meta-analysis. J. Clin. Psychiatry 80 (2). https://doi.org/10.4088/JCP.18r12188.
Xiong, J., Wen, J.L., Pei, G.S., Han, X., He, D.Q., 2023. Effectiveness of internet-based cognitive behavioural therapy for employees with depression: a systematic review and meta-analysis. Int. J. Occup. Saf. Ergon. 29 (1), 268–281. https://doi.org/10.1080/10803548.2022.2043647.
Yang, D., Hur, J.-W., Kwak, Y.B., Choi, S.-W., 2018. A systematic review and meta-analysis of applicability of web-based interventions for individuals with depression and quality of life impairment. Psychiatry Investig. 15 (8), 759–766. https://doi.org/10.30773/pi.2018.03.15.
Zhong, S., Yang, X., Pan, Z., Fan, Y., Chen, Y., Yu, X., Zhou, L., 2023. The usability, feasibility, acceptability, and efficacy of digital mental health services in the COVID-19 pandemic: Scoping review, systematic review, and metaanalysis. JMIR Public Health Surveil. 9, e43730. https://doi.org/10.2196/43730.