Meta-analytic review of online guided self-help interventions for depressive symptoms among college students

Liang Ma a,*, Cong Huang b, c, Rimeng Tao a, Zeshi Cui d, e, Philip Schluter e

a The First Hospital of China Medical University, Shenyang, China
b Department of Sports and Exercise Science, Zhejiang University, Hangzhou, China
c Department of Medicine and Science in Sports and Exercise, Tohoku University Graduate School of Medicine, Sendai, Japan
d School of Pharmacy, China Medical University, Shenyang, China
e School of Health Sciences, University of Canterbury, Christchurch, New Zealand

ARTICLE INFO

Keywords: Internet Guided Depressive symptoms College students Meta-analysis

ABSTRACT

Objective: This meta-analysis examines the effect of online guided self-help interventions for depressive symptoms among college students.

Methods: We searched studies through PubMed, Embase, Web of Science, PsycINFO, and Cochrane Central. Effect estimates were reported as standardized mean differences (SMD) and data were pooled using random-effects models. Subgroup analyses were conducted to investigate the differential effects of these interventions by sample type, level of contact, use of incentive, length of intervention, and program content.

Results: 24 comparisons (n = 3074) deriving from 19 trials were included in the meta-analysis. Intervention participants (n = 1620) indicated significant reductions in depressive symptoms at post-intervention compared to non-active control conditions (n = 1454). The weighted effect size was 0.46 (95% CI: 0.28–0.64), which dropped to 0.36 (95% CI: 0.26–0.45) after an outlier was removed. Subgroup analyses showed that the effects were significant among interventions using both selective and universal samples; among interventions of shorter (<4 weeks), moderate (4–8 weeks), and greater length (≥8 weeks); among interventions with high, moderate, and low levels of contact; among interventions with and without incentive; and among interventions employing cognitive-behavioral therapy (CBT) and third-wave CBT.

Conclusion: This meta-analysis reinforces evidence to support the effectiveness of online guided self-help interventions in reducing depressive symptoms among college students. However, because of the generally variable and limited quality of current evidence, further research applying rigorous methods is needed to confirm and extend the findings of this meta-analysis.

1. Introduction

The college years are characterized by a developmentally challenging transition from adolescence to adulthood, which represents a peak age period for depression onset (Cuijpers et al., 2016; Ibrahim et al., 2013; Zivin et al., 2009). Elevated levels of depressive symptoms are common among college students, leading to considerable impairment and higher risk for depression (Ibrahim et al., 2013). According to the 2018 American College Health Survey, 41.9% of this population felt so depressed that it was difficult to function at least once within the previous year and 18.1% were diagnosed or treated for depression within the last 12 months (American College Health Association, 2018). However, very few seek help despite the fact that many colleges offer counseling services (Eisenberg et al., 2009; Zivin et al., 2009). Research has shown that many students prefer to manage mental health issues alone, due to concerns on confidentiality, a feeling that the problem is too personal, and the fear that no person or service could help (Ebert et al., 2019; Gould et al., 2002). In addition, fear of stigma, lack of time, financial costs, and inconvenient access to care have been identified as significant barriers to seeking help among college students (Gulliver and Bennett, 2015; Hunt and Eisenberg, 2010; Zivin et al., 2009). Because of these important hindrances, innovative interventions that overcome the obstacles are necessary to address depression for the student population.

Online self-help programs may have the potential to improve the delivery of mental health service in colleges. In comparison with the general population, college students are more likely to use the internet...
and to seek health information online (Chiauzzi et al., 2008; Hanauer et al., 2004). Such internet-based interventions are easily accessible and available all the time to participants; do not have to involve a dedicated therapist, allowing participants to stay anonymous without adopting a patient role; and are cost-effective especially when involving some guidance (Donker et al., 2015; Health Quality Ontario, 2019; Mitchell et al., 2021). Since students experiencing mental health problems infrequently seek help from professionals due to various barriers but tend to be comfortable with modern information technologies, online self-help interventions may be of especial interest to the student population. It is possible that, for many students, online self-help interventions may be an acceptable alternative to traditional face-to-face formats. For instance, students who are reluctant to seek face-to-face counseling due to the real or perceived barriers to accessing services may be willing to use an online self-help program that can be accessed privately at a time that is convenient.

Despite the advantages of online self-help interventions, one potential shortcoming of these interventions is low program completion rates (Andrews et al., 2018), which may diminish the effectiveness of an intervention. Multiple meta-analyses examining the efficacy of internet-based interventions suggest that users reap larger benefits from those including provision of human support and guidance (Andersson and Cuijpers, 2009; Baumeister et al., 2014; Heber et al., 2017). It is encouraging that recent meta-analyses suggest that guided self-help interventions for depression can have comparable effects to face-to-face interventions as well as equal adherence (Cuijpers et al., 2010a; van Ballegooijen et al., 2014; Carlbring and Andersson, 2018). In this review, we restricted the focus to guided online self-help interventions, which can be distinguished from other internet-based interventions by the support that is given by health professionals (e.g., public health nurses, psychotherapists) or program coaches to the participants when working through the intervention.

To date, a number of internet-based interventions either guided or unguided have been developed to address depression among the college population. Although associations of these interventions with depression have been evaluated in randomized trials, results of these studies were frequently limited by small sample size, high dropout rates, and inconsistent findings. Only one qualitative review and two meta-analytic reviews have examined the effectiveness of internet-based interventions for depression targeting this population (Davies et al., 2014; Harrer et al., 2019; Lattie and Adkins, 2019). Although the aforementioned meta-analyses showed positive results on depressive symptoms with effect sizes in a small range, all authors combined guided and unguided self-help interventions in their investigations and there was substantial variation in the degree of success for included programs. As stated earlier, guided self-help interventions differ significantly from those unguided in terms of dropout rates and intervention effects, and should therefore be analyzed independently. However, no meta-analysis has specifically focused on effect studies of online guided self-help interventions for college students and examined potential effect moderators and other sources of heterogeneity. Restricting the focus to online guided self-help interventions in this meta-analysis will increase homogeneity, and enable us to assess their independent impact specific to this population.

In addition to examining the overall effect size, meta-analysis also provides opportunities to calculate and compare the separate effect sizes for specific subgroups of interest. Such information would be useful as there are different subgroups forever information needs. Research has shown that intervention programs for depression targeting at-risk individuals (selective and indicated programs) exhibited more support than those delivered universally (universal programs), possibly because high-risk participants are more motivated to engage in the program content and have a greater opportunity to show symptom reduction (Horowitz and Judy, 2006). Knowledge of such differences lends itself well to the suggestion that the effects of online guided self-help interventions for depression among college students may vary across universal, selective and indicated programs. Furthermore, whilst including therapist guidance could be beneficial, the increased cost associated with providing such professional support may prevent colleges with insufficient resources from adopting internet programs. Support has emerged that it is not necessary for the person providing such guidance to be a professionally trained therapist (Robinson et al., 2010; Titov et al., 2010). Thus it would be of value to evaluate whether online self-help interventions with different levels of contact (e.g., therapist, coach, and standardized emails) would produce equal effects for college students. Moreover, although meta-analyses on other problems such as smoking and eating disorder suggested superior effects for longer interventions to very brief interventions (Rooney and Murray, 1996; Stice and Shaw, 2004), larger effects emerged for shorter than longer interventions in a meta-analysis for adolescent depression (Stice et al., 2009). In the college context, it is possible that longer programs require significant time and may not be practical or appealing to college students who commonly have a crowded agenda. This meta-analysis therefore investigated the differential effects of such interventions by length of intervention. In addition, it is unclear whether increasing students’ engagement through providing compensation (e.g., course credit, cash, and raffles) for participating in the program would produce greater intervention effects and it would be helpful for this meta-analysis to assess the difference between interventions with and without such incentive. Finally, intervention content may also affect whether a program produces effects, as elements that form the basis of different strategies have varied widely. Depression intervention programs can employ a broad range of strategies and examining them separately can help inform which strategy would produce greater benefit among the student population.

Therefore, a meta-analytic review was undertaken to assess the effectiveness of online guided self-help interventions for depressive symptoms in college students. It was hypothesized that students assigned to these interventions would show lower depressive symptoms than those assigned to control conditions. A secondary aim was to examine the effectiveness of online guided self-help interventions in different contexts. We expected to find differential effects of these interventions by type of sample, level of contact, use of incentive, length of intervention, and intervention content.

2. Methods

This study was conducted in adherence to the PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions (Moher et al., 2009); see Appendix 1.

2.1. Identification and selection of studies

Studies were identified by searching PubMed, Embase, Web of Science, PsycINFO, and Cochrane Central for papers from inception to September 13 2019, and updated to April 1, 2021. The literature search was constructed around search terms for relevant participants (e.g., college students), search terms for relevant interventions (e.g., internet program), search terms for relevant outcomes (e.g., depression), and search terms for relevant design (e.g., randomized). The search strategy used in PubMed is displayed in Appendix 2, which was adapted for other databases as necessary. Reference lists from relevant reviews and retrieved studies were reviewed and searched to identify additional records.

Initial records identified from electronic database searching, relevant reviews, and retrieved studies were imported into Endnote X7. After removal of duplicates, titles and abstracts were screened by the primary researcher (LM) and full-text articles for studies deemed potentially relevant were retrieved for further assessment by two independent researchers (LM and CH). Disagreements between the two researchers were resolved through discussion and consensus, and a third researcher (ZC) who was consulted if agreement between the two researchers could
not be reached.

We included (a) randomized controlled trials (RCTs) that compared (b) an online guided self-help intervention (which was considered eligible when the intervention used the Internet as delivery mode via computer, laptop, tablet, or other device, and included some degree of personal guidance or at least regular e-mail or other non-technical contacts with a therapist or coach) to (c) a non-active control condition including wait-list, assessment-only, and usual-care (in which participants were permitted to continue using the care as usual such as mental health counseling services) in terms of effects on (d) depressive symptoms as a primary or secondary (which was regarded as eligible only when the primary outcome was distress, anxiety or stress) outcome using (e) validated measures among (f) non-clinical samples at a (g) higher education facility (e.g., university, college, tertiary, or comparable post-secondary education).

2.2. Data extraction

Information was extracted using a piloted abstraction form for each study based on: (a) the characteristics of the trial including country and year of publication, sample size, outcome measures, and type of control; (b) the characteristics of the intervention including intervention content, follow-up in weeks, level of contact, use of incentive, and involvement of group technique; and (c) the characteristics of participants including student sample, recruitment strategy and setting, type of sample, mean age, and percentage of females.

Type of sample was categorized as universal if the intervention was administered to all members of a population without explicit screening procedures on symptoms or risk factors of depression; selective if participants were included on the basis of endorsing a known risk factor of depression; and indicated if the intervention was delivered to individuals reporting elevated but subclinical levels of depressive symptoms (Mrazek and Hagerty, 1994). Level of contact was categorized as high if qualified psychotherapists monitored the intervention and provided support directly; moderate if the support was provided by research team members or program coaches such as trained students; and low if only standardized reminder emails were used to deliver the support. Length of intervention was categorized as brief (≤4 weeks), moderate (4–8 weeks), and long (>8 weeks) following the categorization of a recent meta-analysis (Harrer et al., 2019). Use of incentive was categorized as yes if students were rewarded by cash, course credit, or other forms of compensation for their engagement in the program; and no if incentive was not used. Intervention content was categorized as CBT if the intervention focused on cognitive restructuring such as identifying and changing negative cognition as well as behavioral activation that emphasizes increasing pleasant activities (Beck, 1979); third-wave CBT if the intervention relied on components based on mindfulness, meditation, and/or acceptance principles (Ost, 2008) such as mindfulness-based therapy (MBT; Kabat-zinn, 2003) and acceptance and commitment therapy (ACT; Hayes, 2004); and others if the intervention utilized additional strategies (e.g., physical activity). Involvement of a group technique was indicated if the intervention included group-based components, such as live videoconferences. Recruitment strategy was categorized as online if the participants were recruited primarily via online advertisements, websites, or circular e-mails; and on-site if recruited through announcements in classes or other institutions. Study setting was categorized as campus if the participants were recruited from specified colleges; and public if recruited from the general population.

2.3. Risk of bias assessment

The methodological quality was determined using the Cochrane Collaboration’s Risk-of-Bias Tool 2 (Sterne et al., 2019), which considers risk of bias across five domains: (1) the randomization process (e.g., adequate generation of random sequence and concealment of allocation to conditions); (2) deviations from the intended interventions (e.g., blinding of participants, carers and trial personnel); (3) missing outcome data (e.g., attrition rate < 10% or use of intention-to-treat analysis); (4) measurement of outcome (e.g., blinding of outcome assessors); and (5) selection of the reported results (e.g., published trial protocols). The risk of bias for each domain was scored as low, moderate, or high. Because we focused on trials that compared online guided self-help interventions with non-active controls such as wait-list, it was difficult in these contexts to ensure that the participants and people delivering the interventions were unaware of the interventions assignment. Therefore, we omitted the domain assessing the risk of bias due to deviations from the intended interventions. The risk of bias assessment was performed independently by two authors (LM and RT), with a third author (IZ) engaged in case of disagreement.

2.4. Statistical analysis

Intervention effects were defined as the difference of mean change scores in depressive symptoms between intervention and control conditions at post-intervention. We calculated the effect size for each study with standardized mean differences (SMD) and 95% confidence intervals (CI). Because a number of studies had small sample sizes, we corrected the effect sizes for small sample bias using Hedges’ g (Hedges and Olkin, 1985): $g = \frac{M_{\text{Intervention}} - M_{\text{Control}}}{SD_{\text{Pool}}} \times k$, where the pooled standard deviation (at pretest) was defined as $SD_{\text{Pool}} = \sqrt{\frac{N_{\text{Intervention}} - 1}{N_{\text{Control}} - 1} SD_{\text{Intervention}}^2 + \frac{N_{\text{Intervention}}}{N_{\text{Control}}^2} (N_{\text{Intervention}} - 1) SD_{\text{Control}}^2}$ and the correction factor was defined as $k = 1 - \frac{3}{4 N_{\text{Intervention}} + N_{\text{Control}} - 3}$.

We used random-effects models and inverse variance weighted methods to evaluate the pooled effect sizes (DerSimonian and Laird, 1986). Effect sizes of 0.5 or lower were considered as small, effect sizes of 0.5–0.8 were interpreted as moderate, and effect sizes higher than 0.8 were regarded as large (Cohen, 2013).

In some studies (namely: Ahmad et al., 2020; Ellis et al., 2011; Gibbel, 2010; Morris et al., 2016; Sethi et al., 2010), multiple intervention arms were included, resulting in two comparisons between an online guided self-help intervention and the control condition. We calculated the effect size for each individual comparison and included both comparisons in the overall meta-analysis. Because such comparisons are dependent, and may distort the pooled effect sizes by artificially reducing heterogeneity (Borenstein et al., 2011), we conducted a series of sensitivity analyses to assess the influence of studies with multiple comparisons on the pooled effect size. First, we recalculated the pooled effect size by including only the comparison with larger effect size from that study. Second, we recalculated the pooled effect size by including only the comparison with the smaller effect size from that study. Finally, we recalculated the pooled effect size by combining the effects of both intervention groups in that study to create a single comparison before being pooled in the overall meta-analysis.

We evaluated the impact of individual studies on the pooled effect sizes by conducting additional influence analyses using a holdout approach whereby each individual study was sequentially omitted from the eligible study pool and analyzed.

Publication bias was assessed using Egger regression symmetry tests, which indicate publication bias if asymmetry is suggested. We also performed the Duval-Tweedie trim and fill procedure, with an adjusted effect size being produced to account for missing studies that lead to publication bias (Duval and Tweedie, 2000). Heterogeneity was assessed using the I$^2$-statistic that was distinguished as low, moderate, and high with values of 25%, 50%, and 75% (Higgins et al., 2003). Subgroup analyses were performed to explore potential sources of heterogeneity. We calculated the pooled effect sizes for subgroups with enough comparisons available (n ≥ 3), including type of sample (universal or selective or indicated), level of contact (high or moderate or low), use of incentive (yes or no), length of...
intervention (brief or moderate or long), and intervention content (CBT or third-wave). All the analyses were performed with Stata release 12 (StataCorp) and \( p \leq 0.05 \) was considered statistically significant.

3. Results

3.1. Selection and inclusion of studies

A total of 3382 records were initially identified from database searching and reference review. After a careful assessment of the identified citations with the eligibility criteria, 19 studies were included in the current review. A PRISMA flowchart summarizing the process for inclusion of studies is presented in Fig. 1.

3.2. Descriptive characteristics of included studies

Descriptive characteristics of the included studies are presented in Table 1. All studies were conducted after 2010, primarily in western countries. In total, the studies included 3074 participants with sample sizes ranging from 38 (Sethi et al., 2010) to 1162 (Viskovich & Pakenham, 2020). All studies but one (Cook et al., 2019) recruited the participants from specified colleges. The participants were primarily recruited online (\( n = 15 \)), whilst the remaining 4 studies used on-site recruitment strategies. The mean age of participants was 22.9 years and the average proportion of females was 73.4%. Of the 19 studies, ten studies used traditional CBT, eight studies used third-wave CBT, and one study used physical activity. There were four studies including high levels of contact, 13 studies providing moderate levels of contact, and 2 studies involving low levels of contact throughout the intervention. The length of intervention ranged from 3 (Ellis et al., 2011; Levin, 2014; Sethi et al., 2010) to 12 weeks (Cook et al., 2019), including 6 brief interventions, 8 moderate interventions, and 5 long interventions. Eight were categorized as universal interventions, nine were categorized as selective interventions, and 2 were categorized as indicated interventions. Eleven studies used incentive to promote students’ engagement in the program whilst 8 studies provided no such compensation. The interventions were predominantly administered individually, with three studies also involving group technique. The most commonly used measure for depressive symptoms was DASS (Depression, Anxiety, and Stress Scales; \( n = 6 \)), PHQ (Patient Health Questionnaire; \( n = 6 \)), and BDI (Beck Depression Inventory; \( n = 5 \)). The majority (\( n = 14 \)) of studies used a wait-list control condition, and the other studies used an assessment-only (\( n = 3 \)) or usual-care (\( n = 2 \)) control condition. For studies providing data on follow-up assessments (\( n = 7 \)), support was shown in all these studies that the intervention effects were maintained at up to 12 months follow-up.

3.3. Risk of bias assessment

Table 2 presents a summary of risk of bias assessment for each of the

![Fig. 1. PRISMA flowchart.](image-url)
| Author                      | Student sample (n)                                                                 | Recruit | Setting     | Age (mean) | Female % | Sample type | Online intervention (n) | Control (n) | Contact level | Incentive | Group format | Symptom measure | Country   |
|-----------------------------|-----------------------------------------------------------------------------------|---------|-------------|------------|----------|-------------|-------------------------|-------------|---------------|-----------|--------------|-----------------|-----------|
| Day et al., 2013            | Students with mild to moderate levels of anxiety, depression or stress: DASS ≥ 8, 10, 15 (66) | Online   | Campus      | 23.6       | 89       | S           | CBT (33)                | WL (33)     | Trained students (M) | No        | --           | DASS CAN       |           |
| Ellis et al., 2011          | Students with elevated scores on the Kessler Psychological Distress Scale: K10 < 30 (39) | On-site  | Campus      | 19.7       | 77       | S           | Online CBT; The MoodGYM intervention (13); Online Peer Support: The MoodGarden intervention (13) | AO (13)     | Researcher (M) | Yes       | --           | DASS AUS       |           |
| Eastis et al., 2018         | General (156)                                                                     | Online   | Campus      | 25         | 79       | U           | Acceptance-based behavioral intervention (78); iCBT (Moodgym) (24); Spiritual intervention (19) | WL (78)     | Doctoral candidate in clinical psychology (M); Research team member (M) | 4         | Yes          | DASS USA       |           |
| Gibbel, 2010                | Students with mildly and moderately depressed undergraduates: CES-D ≥ 10 and <25 (65) | Online   | NS          | 83         | 83       | I           | CBT (30)                | WL (75)     | Trained students (M) | No        | --           | BDI GER        |           |
| Harrer et al., 2018         | Students with elevated levels of stress: PSQ-4 ≥ 8 (150) First-year students (76) | Online   | Campus      | 24.1       | 75       | S           | iCBT for stress (StudiCare Stress) (75); ACT (37) | WL (39)     | Regular contact (M) | Yes       | --           | DASS USA       |           |
| Levin, 2014                 | General (79)                                                                      | Online   | Campus      | 21         | 66       | U           | ACT (40)                | WL (39)     | Research assistants (M) | 4         | Yes          | CCAPS-D USA    |           |
| Levin et al., 2017          | Students with receiving mental health counseling (47) General (128)               | Online   | Campus      | 20.5       | 67       | U           | CBT programs: “Insomnia Relief” (48) and “Anxiety Relief” (43) | WL (47)     | Standardized emails (L) | 6         | Yes          | BDI UK         |           |
| Mullin et al., 2015         | Students with self-identified as experiencing symptoms of anxiety or depression (55) | Online   | Campus      | 27.9       | 64       | S           | CBT (30)                | WL (23)     | Psychologist (H) | 6, 12      | Yes          | PHQ AUS        |           |
| Rasinen et al., 2016        | Students with self-reporting as experiencing some form of psychological distress (68) | Online   | Campus      | 24         | 85       | S           | ACT (33)                | WL (35)     | Trained psychology students (M) | 7, 48      | No           | BDI IRL        |           |
| Richards et al., 2016       | Students with self-reported GAD symptoms: GAD-7 ≥ 10 (127)                         | Online   | Campus      | 23.8       | 77       | S           | CBT (70)                | WL (67)     | Psychologists (H) | 6         | No           | BDI USA        |           |
| Sethi et al., 2010          | First-year students with low to moderate levels of depression and/or anxiety (38) | On-site  | Campus      | 19.5       | 66       | S           | iCBT (Moodgym) (9); Combined face-to-face + online CBT (9) | AO (10)     | Psychologists (H) | 3         | Yes          | DASS AUS       |           |
| Viskovich & Pakenham, 2020  | Undergraduates (154)                                                              | Online   | Campus      | 23.1       | 76       | U           | Mindfulness virtual community (76) | WL (78)     | Moderators (M) | Yes       | 8            | LVC CAN        |           |
| Ritvo et al., 2021          | Students with elevated repetitive thoughts: PSWQ ≥ 50; RRS ≥ 40 (235)              | Online   | Public      | 20.4       | 83       | S           | CBT (82)                | UC (77)     | Clinician (H) | 12        | No           | PHQ UK         |           |
| Cook et al., 2019           | Students with mild to moderately severe depressive symptoms: PHQ ≥ 10 and <19 (214) | Online   | Campus      | 22.2       | 71       | I           | CBT (107)               | WL (107)    | Postgraduate student in clinical psychology (M) | 7, 12      | No           | PHQ UK         |           |
| Salamanca-Sanabria et al., 2020 | Students with mild to moderately severe depressive symptoms: PSQ-4 ≥ 8 (235) | Online   | Campus      | 22.2       | 71       | I           | CBT (107)               | WL (107)    | Postgraduate student in clinical psychology (M) | 7, 12      | No           | PHQ UK         |           |
| Ahmad et al., 2020          | Undergraduates (119)                                                              | Online   | Campus      | 24.8       | 75       | U           | Full mindfulness virtual community (40); partial mindfulness virtual community (39) | WL (40)     | Moderators (M) | 8         | Yes          | LVC PHQ CAN    |           |
| El Morr et al., 2020        | Undergraduates (160)                                                              | Online   | Campus      | 22.6       | 79       | U           | Mindfulness virtual community (79) | WL (80)     | Moderators (M) | 8         | Yes          | LVC PHQ CAN    |           |

Note. U = universal; S = selective; I = indicated; M = moderate; H = high; WL = wait-list; UC = usual-care; AO = assessment-only; LVC = live video conference. NS = not specified; BDI = Beck Depression Inventory; DASS = Depression, Anxiety, and Stress Scales; PHQ = Patient Health Questionnaire; CES-D = Centre for Epidemiologic Studies-Depression; CCAPS-D = Counseling Center Assessment of Psychological Symptoms-Depression.
Table 2
Risk of bias assessment.

| Author                       | The randomization process | Missing outcome data | Measurement of the outcome | Selection of the reported result | Overall     |
|------------------------------|----------------------------|----------------------|----------------------------|---------------------------------|-------------|
| Day et al., 2013             | +                         | +                    | +                          | –                               | Moderate    |
| Ellis et al., 2011           | –                         | –                    | +                          | –                               | High        |
| Eustis et al., 2018          | –                         | –                    | +                          | –                               | High        |
| Gibbel, 2010                 | –                         | –                    | –                          | –                               | High        |
| Harrer et al., 2018          | +                         | +                    | +                          | –                               | Moderate    |
| Levin, 2014                  | –                         | +                    | +                          | –                               | Moderate    |
| Levin et al., 2017           | –                         | +                    | +                          | –                               | Moderate    |
| Mailey et al., 2010          | –                         | +                    | +                          | –                               | High        |
| Morris et al., 2016          | –                         | +                    | +                          | –                               | Moderate    |
| Mullin et al., 2015          | –                         | +                    | +                          | –                               | Moderate    |
| Rasanen et al., 2016         | –                         | +                    | –                          | +                               | High        |
| Richards et al., 2016        | +                         | +                    | +                          | +                               | Low         |
| Sethi et al., 2010           | –                         | +                    | –                          | –                               | High        |
| Viskovich & Pakenham, 2020   | –                         | +                    | +                          | –                               | Moderate    |
| Ritvo et al., 2021           | +                         | +                    | +                          | –                               | Moderate    |
| Cook et al., 2019            | +                         | +                    | –                          | +                               | Moderate    |
| Salamanca-Sanabria et al., 2020 | +                  | –                    | +                          | –                               | High        |
| Ahmad et al., 2020           | +                         | +                    | +                          | –                               | Moderate    |
| El Morr et al., 2020         | +                         | +                    | +                          | –                               | Moderate    |

Note. + = lower risk of bias; – = higher risk of bias; ~ = some concern.

Studies without outlier

| Author                        | SMD (95% CI)     | Weight % |
|-------------------------------|------------------|----------|
| Ahmad 2020 (Full Mindfulness Virtual Community) | 0.44 (−0.10, 0.98) | 4.04     |
| Ahmad 2020 (Partial Mindfulness Virtual Community) | 0.41 (−0.14, 0.95) | 4.04     |
| Cook 2019                     | 0.18 (−0.14, 0.49) | 5.36     |
| Day 2013                      | 0.40 (−0.09, 0.89) | 4.35     |
| Ellis 2011 (MoodGYM)          | 0.45 (−0.53, 1.43) | 2.21     |
| Ellis 2011 (MoodGarden)       | 0.83 (−0.13, 1.80) | 2.27     |
| Eustis 2018                   | 0.52 (0.20, 0.84)  | 5.32     |
| Gibbel 2010 (MoodGYM)         | 0.06 (−0.69, 0.80) | 3.07     |
| Gibbel 2010 (Here Comes the Sun) | 0.78 (0.04, 1.52)  | 3.08     |
| Harrer 2018                   | 0.67 (0.34, 1.00)  | 5.26     |
| Leven 2014                    | 0.34 (−0.11, 0.80) | 4.55     |
| Levin 2017                    | 0.18 (−0.26, 0.62) | 4.61     |
| Mailey 2010                   | 0.11 (−0.44, 0.66) | 4.01     |
| Morr 2020                     | 0.44 (0.13, 0.75)  | 5.35     |
| Morris 2016 (Anxiety Relife)  | 0.29 (−0.21, 0.79) | 4.27     |
| Morris 2016 (Insomnia Relife) | 0.23 (−0.27, 0.73) | 4.29     |
| Mullin 2015                   | 0.64 (0.09, 1.20)  | 3.96     |
| Rasanen 2016                  | 0.69 (0.20, 1.18)  | 4.34     |
| Richards 2016                 | 0.59 (0.25, 0.94)  | 5.19     |
| Ritvo 2021                    | 0.07 (−0.25, 0.39) | 3.33     |
| Sethi 2010 (Combined face-to-face/online CBT) | 1.02 (−0.16, 2.20) | 1.71     |
| Sethi 2010 (CBT)              | −0.42 (−1.53, 0.69) | 1.87     |
| Viskovich 2019                | 0.23 (0.12, 0.35)  | 6.21     |
| Subtotal (I²-squared = 13.2%, p = 0.281) | 0.36 (0.26, 0.45) | 94.68    |

Outlier

| Author                        | SMD (95% CI)     | Weight % |
|-------------------------------|------------------|----------|
| Salamanca-Sanabria 2020       | 1.79 (1.47, 2.11) | 5.32     |
| Subtotal (I²-squared = .%, p = .) | 1.79 (1.47, 2.11) | 5.32     |
| Levin 2014                    | 0.46 (0.28, 0.64) | 100      |

NOTE: Weights are from random effects analysis

Fig. 2. Forest plot for the meta-analysis of online guided self-help interventions on depressive symptoms.
included studies. Overall, the risk of bias was variable across studies: only one study fully met all the remaining four domains and was categorized as low risk; eight studies showed high risk for at least one domain and was categorized high risk; and 10 studies were categorized as moderate risk as concerns emerged for one or more domains.

### 3.4. Meta-analytic results

We were able to calculate the effect sizes indicating the difference between online guided self-help interventions and control conditions in 24 comparisons that derived from the 19 studies. Forest plot for the meta-analysis results on depressive symptoms is presented Fig. 2. At posttest, 15 comparisons (63%) produced non-significant effect sizes (the 95% Cs overlapped zero); 8 comparisons (33%) produced significant effect sizes in small-to-moderate range, and 1 comparison (4%) categorized as low risk; eight studies showed high risk for at least one comparison (4%) produced a significant and large effect size.

The pooled effect size for depressive symptoms at posttest was \( g = 0.46 \) (95% CI: 0.28–0.64), which was considered small but significant (the 95% CI did not overlap zero). Heterogeneity for depressive symptoms was high and significant (\( I^2 = 77.6%; p < 0.001 \)). Visual inspection of the screen plot indicated that the study by Salamanca-Sanabria et al. (2020) (which targeted students with mild to moderately severe depressive symptoms) produced the largest effect size and was discrepant with the overall distribution. After removal of this potential outlier, the pooled effect size dropped to 0.36 (95% CI: 0.26–0.45), and the heterogeneity turned non-significant and was reduced to a low level (\( I^2 = 13.2%; \ p = 0.28 \)). Because of its considerable influence on the pooled effect size, we excluded this study from subsequent analyses.

In the sensitivity analyses of studies with multiple comparisons, the resulting effect size was \( g = 0.38 \) (95% CI: 0.27–0.48) when comparisons with the smaller effect size from the same study were removed; the resulting effect size was \( g = 0.35 \) (95% CI: 0.25–0.45) on exclusion of comparisons with the larger effect size; and the resulting effect size was \( g = 0.35 \) (95% CI: 0.26–0.45) when combined effects of multiple comparisons were considered in the meta-analysis. Heterogeneity was non-significant and remained at low levels (all \( I^2 \leq 25\% \); all \( p > 0.05 \)) in these analyses. These results suggested no substantial changes in the pooled effect sizes compared with the ones found in the overall analysis.

In the influence analyses of individual studies, the pooled effect size ranged from 0.31 (95% CI: 0.24–0.39) on exclusion of the study by Harrer et al. (2018) to 0.40 (95% CI: 0.30–0.50) after removal of the study by Viskovich and Pakenham (2020). There was no indication of significant heterogeneity (both \( I^2 = 0 \); both \( p > 0.05 \)) in the analyses in which either of the studies was omitted. Overall, these analyses indicated that removal of individual studies resulted in no substantial changes in the pooled effect sizes.

We further assessed the impact of 5 studies in which wait-list control condition was not used, by omitting these studies from the analyses. The heterogeneity was non-significant and low (\( I^2 = 20.5%; p = 0.23 \)), and the resulting effect size was \( g = 0.38 \) (95% CI: 0.27–0.48), which did not change substantially compared with the overall analysis. Similar analyses were performed to evaluate the influence of three studies that involved group-based components. The resulting effect size was again \( g = 0.38 \) (95% CI: 0.27–0.48) with a non-significant and low level of heterogeneity (\( I^2 = 18.3%; p = 0.23 \)), which was also comparable to the effect size found in the overall meta-analysis.

There was no evidence for asymmetry of the funnel plot in the overall meta-analysis as Egger’s test was non-significant (\( p = 0.09 \)), which indicated low levels of publication bias. Under the trim and fill procedure, the adjusted effect size was stable (\( g = 0.35; 95\% \ CI: 0.26–0.43 \), with three imputed studies being incorporated into the model.

Results for the subgroup analyses were available in Table 3. None of the hypothesized variables accounting for any significant between-subgroup differences (all \( p > 0.05 \)). We focused on analyses evaluating whether the effects of these interventions were significant in specific subgroups of interest. Significant effects were replicated across interventions using both universal (\( g = 0.27; 95\% \ CI: 0.18–0.36 \)) and selective approaches (\( g = 0.47; 95\% \ CI: 0.30–0.64 \)), interventions of shorter (\( g = 0.29; 95\% \ CI: 0.16–0.42 \), moderate (\( g = 0.52; 95\% \ CI: 0.37–0.68 \)), and greater length (\( g = 0.25; 95\% \ CI: 0.09–0.41 \)), interventions of high (\( g = 0.42; 95\% \ CI: 0.10–0.74 \), moderate (\( g = 0.40; 95\% \ CI: 0.29–0.52 \)), and low levels of contact (\( g = 0.24; 95\% \ CI: 0.13–0.34 \), interventions with (\( g = 0.36; 95\% \ CI: 0.23–0.48 \)) and without incentive (\( g = 0.39; 95\% \ CI: 0.21–0.57 \)); and interventions employing traditional (\( g = 0.42; 95\% \ CI: 0.27–0.57 \)) and third-wave CBT (\( g = 0.32; 95\% \ CI: 0.21–0.44 \)). Heterogeneity was low and non-significant (all \( I^2 < 25\% \); all \( p > 0.05 \)) in these subgroup analyses.

We also conducted a subgroup analysis to examine the potential influence of study quality on the intervention effects. The effect size based on comparisons categorized as high risk of bias was \( g = 0.45 \) (95% CI: 0.27–0.63); and the effect size based on comparisons categorized as moderate risk of bias was \( g = 0.29 \) (95% CI: 0.20–0.39). Heterogeneity stayed low and non-significant in these analyses (both \( I^2 < 10\% \); both \( p > 0.05 \)), and no significant difference (\( p = 0.52 \)) was produced between these subgroups, suggesting the results are robust despite the inclusion of studies with mixed quality.

We also attempted to examine the long-term effectiveness of online guided self-help interventions for depressive symptoms among college students. However, owing to the small number of studies (\( n = 7 \)) with variable follow-up intervals (which ranged from 6 to 60 weeks) and

### Table 3

| Subgroup analyses by hypothesized variables. | Effect sizes | P value between subgroups |
|:----------|:-------------|:--------------------------|
| **Type of sample** | **Hedge’s g (95% CI)** | **Heterogeneity** |
| Universal | 0.27 (0.18, 0.36) | \( I^2 = 0, p = 0.71 \) |
| Selective | 0.47 (0.30, 0.64) | \( I^2 = 16.9%, p = 0.28 \) |
| Length of intervention | | |
| Brief | 0.29 (0.16, 0.42) | \( I^2 = 61.1%, p = 0.38 \) |
| Moderate | 0.52 (0.37, 0.68) | \( I^2 = 0, p = 0.66 \) |
| Long | 0.25 (0.09, 0.41) | \( I^2 = 0, p = 0.56 \) |
| Level of contact | | |
| High | 0.42 (0.10, 0.74) | \( I^2 = 43.8%, p = 0.13 \) |
| Moderate | 0.40 (0.29, 0.52) | \( I^2 = 0, p = 0.54 \) |
| Low | 0.24 (0.03, 0.34) | \( I^2 = 0, p = 0.98 \) |
| Use of incentive | | |
| Yes | 0.36 (0.23, 0.48) | \( I^2 = 0, p = 0.64 \) |
| No | 0.39 (0.21, 0.57) | \( I^2 = 52.0%, p = 0.05 \) |
| Intervention content | | |
| CBT | 0.42 (0.27, 0.57) | \( I^2 = 6.6%, p = 0.38 \) |
| Third-wave | 0.32 (0.21, 0.44) | \( I^2 = 15.2%, p = 0.30 \) |
insufficient follow-up data (e.g., follow-up assessments were only conducted within the intervention condition), we were unable to calculate the follow-up effect sizes indicating the difference between online guided self-help interventions and control conditions in the longer term.

4. Discussion

This meta-analysis examined 19 studies using online guided self-help interventions to address depressive symptoms for a total of 3074 college students. We found that students who participated in these interventions reported lower levels of depressive symptoms compared to those who received no intervention. The effect size found in this meta-analysis is modest but compares favorably to the one reported in Harrer’s recent meta-analysis, which combined guided and unguided internet-based interventions targeting college students. Such strengthening effects may be attributable to the focus on guided self-help interventions in the current meta-analysis as stronger effects were found for guided versus unguided online interventions in a range of other target groups (Heber et al., 2017; Richards and Richardson, 2012; Karyotaki et al., 2021).

Our analyses detected no significant difference for online guided self-help interventions with different levels of contact (e.g., therapist, coach, or standardized emails), suggesting that such interventions may produce equivalent effects. In support, two other meta-analyses have reported similar findings with regard to the equivalent efficacy of therapist versus non-therapist supported online interventions for depression among the general population (Richards and Richardson, 2012; Newby et al., 2016). However, this failure to find a significant difference may be due to a lack of power, since our analyses were based on a relatively small number of studies. Thus, further studies are needed to determine whether internet-based interventions involving different levels of contact would produce equal effects for addressing depressive symptoms among college students.

Significant effects were exhibited for online guided self-help interventions among both universal and selective samples, with albeit non-significantly greater effects for selective interventions. In contrast to multiple meta-analyses of different depression intervention programs supporting superior effects among selective to universal samples (Horwitz and Judy, 2006; Merry et al., 2004; Stice et al., 2009), our analyses point at the potential of equivalent effects for online guided self-help interventions among these samples in the college context. It is possible that online guided self-help interventions are of greater appeal to college students especially when administered universally. However, drawing definitive conclusions based on current evidence is premature and more research is needed.

The effect size of online guided self-help interventions with and without incentive was both significant and comparable in our analyses. Although providing incentive might attract more students to participate, it would be helpful to revisit the question of whether providing incentive truly affects the intervention effects.

Our analyses indicated no significant difference between online guided self-help interventions of moderate length and those briefer or longer interventions. Such findings did not replicate the results of other meta-analyses which suggested significantly stronger effects for interventions of moderate length (Harrer et al., 2019; Heber et al., 2017; Richards and Richardson, 2012). Theoretically, longer interventions enable participants to more sufficiently reflect on the program content and practice the attitudinal and behavioral change skills. However, there could be an upper limit for this occasion as extremely long programs may not appeal to youth, leading to increased dropout rates and attenuating intervention effects. Thus, further trials are needed to establish whether intervention length is associated with improved outcomes, and, if so, the optimal duration of such interventions.

This meta-analysis also examined the difference between traditional and third-wave CBT. In Harrer et al.’s meta-analysis, the authors found significantly higher effects on depressive symptoms among college students for programs based on CBT principles than other programs. However, they used a broader definition of CBT that combined both traditional and third-wave CBT in their analyses, and did not compare the effects between different CBT strategies. Thus, this meta-analysis provided support for Harrer et al.’s approach of combining traditional and third-wave CBT as we found no significant difference between these strategies. As research accumulates, it would be worth further exploring whether providing guidance for different intervention strategies would produce more efficacious effects. For instance, it is possible that third-wave CBT may be delivered more efficaciously if some guidance is provided, whereas increasing guidance may not lead to a significant influence on the effects of traditional CBT. Future research comparing directly traditional and third-wave CBT with the same level of guidance is also needed before we can draw definitive conclusions about the superiority of these different intervention strategies.

This meta-analysis was not pre-registered and has several limitations. First, only English language papers prior to April 1, 2021 were identified, which may omit potentially germane studies (e.g., more current or non-English studies) from inclusion within the meta-analysis. Second, this meta-analysis focused on inactive control conditions and the quality of many studies included in the meta-analysis was variable, which may produce an inflated estimate for the efficacy of these interventions (Cuipers et al., 2010b; Sterne et al., 2000; Thornton and Lee, 2000). Third, this meta-analysis was limited to examine the follow-up effects of online guided self-help interventions as many studies did not report longer-term outcomes and the studies that did report these data employed various follow-up intervals. Forth, this meta-analysis included a range of subgroup contrasts that were based on a small number of studies (n ≤ 10) and many of these studies had small samples, limiting our analyses to detect significant between-subgroup differences.

This meta-analysis has several implications. First, it reinforces that online guided self-help interventions are associated with modest reductions in depressive symptoms among college students. Second, it underlines a range of research gaps. For instance, many of the studies used universal or selective samples with predominant females and provided data solely at the conclusion of intervention, whilst few studies were administered to indicated samples, focused on male students, or conducted follow-up assessments. Third, it provides preliminary information on the mechanisms of how online guided self-help interventions operate among college students by examining a range of hypothesized variables. Experimentally manipulating these variables such as level of contact, length of intervention, and use of incentive would be essential for future research to confirm such ostensive relations.

These results suggest that online guided self-help interventions are likely to be a promising approach for addressing depressive symptoms among college students. However, because of the generally variable and limited quality of current evidence, further research applying rigorous methods is needed to confirm and extend the findings of this meta-analysis.

Declaration of competing interest

All the authors declare that they have no conflicts of interest.
Appendix 1

### PRISMA checklist

| Section/topic                  | # | Checklist item                                                                 | Reported on page # |
|-------------------------------|---|--------------------------------------------------------------------------------|-------------------|
| Title                          | 1 | Identify the report as a systematic review, meta-analysis, or both.              | 2                 |
| Abstract                       | 2 | Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and analysis methods; results; limitations; conclusions and implications of key findings; systematic review registration number. | 1                 |
| Introduction                   | 3 | Describe the rationale for the review in the context of what is already known.   | 2–6               |
| Rationale                      | 3 | Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS). | 6–7               |
| Methods                        | 5 | Indicate if a review protocol exists, and if where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number. | N/A               |
| Eligibility criteria           | 6 | Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale. | 7                 |
| Information sources            | 7 | Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched. | 7                 |
| Search                         | 8 | Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated. | 7, 25             |
| Study selection                | 9 | State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis). | 8                 |
| Data collection process        | 10| Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators. | 8                 |
| Data items                     | 11| List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made. | 8–9               |
| Risk of bias in individual studies | 12| Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis. | 9–10              |
| Summary measures               | 13| State the principal summary measures (e.g., risk ratio, difference in means).     | 10                |
| Synthesis of results           | 14| Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I²) for each meta-analysis. | 11–12             |
| Risk of bias across studies    | 15| Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies). | 11                |
| Additional analyses            | 16| Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified. | 11–12             |
| Results                        | 17| Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram. | 12                |
| Study characteristics          | 18| For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations. | 37–38             |
| Risk of bias within studies    | 19| Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12). | 13, 39            |
| Results of individual studies  | 20| For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot. | Fig. 2            |
| Synthesis of results           | 21| Present results of each meta-analysis done, including confidence intervals and measures of consistency. | 14–16             |
| Risk of bias across studies    | 22| Present results of any assessment of risk of bias across studies (see item 15). | 14–16             |
| Additional analysis            | 23| Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see item 16]). | 14–17             |
| Discussion                     | 24| Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers). | 17–19             |
| Limitations                    | 25| Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias). | 20                |
| Conclusions                    | 26| Provide a general interpretation of the results in the context of other evidence, and implications for future research. | 21                |
| Funding                        | 27| Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review. | N/A               |

Appendix 2

The search strategy used in PubMed for this meta-analysis is presented as follows:

1. Terms related to relevant participants: “college students” OR “university students” OR “post-secondary students” OR “undergraduate students” OR “graduate students”
2. Terms related to relevant interventions: (internet OR online OR web) AND (intervention OR program OR training OR prevention)
3. Terms related to relevant outcomes: depress* OR gradient*
4. Terms related to relevant design: random* AND control*

Search strategy: #1 AND #2 AND #3 AND #4.
Salamanca-Sanabria, A., Richards, D., Timulak, L., Connell, S., Mojica Perilla, M., Parra-Villa, Y., Castro-Camacho, L., 2020. A culturally adapted cognitive behavioral Internet-delivered intervention for depressive symptoms: randomized controlled trial. JMIR Mental Health 7 (1), e13392.

Sethi, S., Campbell, A.J., Ellis, L.A., 2010. The use of computerized self-help packages to treat adolescent depression and anxiety. J. Technol. Hum. Serv. 28 (3), 144–160.

Sterne, J.A., Gavaghan, D., Egger, M., 2000. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J. Clin. Epidemiol. 53, 1119–1129.

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. Br. Med. J. 366, i4898.

Stice, E., Shaw, H., 2004. Eating disorder prevention programs: a meta-analytic review. Psychol. Bull. 130, 206–227.

Stice, E., Shaw, H., Bohon, C., Marti, C.N., Rohde, P., 2009. A meta-analytic review of depression prevention programs for children and adolescents: factors that predict magnitude of intervention effects. J. Consult. Clin. Psychol. 77, 486.

Thornton, A., Lee, P., 2000. Publication bias in meta-analysis: its causes and consequences. J. Clin. Epidemiol. 53, 207–216.

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

Viskovich, S., Pakenham, K.I., 2020. Randomized controlled trial of a web-based acceptance and commitment therapy (ACT) program to promote mental health in university students. J. Clin. Psychol. 76 (6), 929–951.

Zivin, K., Eisenberg, D., Golub, S.E., 2009. Persistence of mental health problems and needs in a college student population. J. Affect. Disord. 117, 180–185.