

Review

Conversational Agent Interventions for Mental Health Problems: Systematic Review and Meta-analysis of Randomized Controlled Trials

Yuhao He^{1,2}, BE, MEd; Li Yang^{1,2}, MEd, PsyD; Chunlian Qian^{1,2}, BS, MEd; Tong Li^{1,2}, BS, MEd; Zhengyuan Su^{1,2}, BS, MEd; Qiang Zhang³, BEd, MEd; Xiangqing Hou^{1,2}, MEd, PhD

¹Institute of Applied Psychology, College of Education, Tianjin University, Tianjin, China

²Laboratory of Suicidology, Tianjin Municipal Education Commission, Tianjin, China

³Shenzhen School, Sun Yat-sen University, Shenzhen, China

Corresponding Author:

Yuhao He, BE, MEd

Institute of Applied Psychology, College of Education

Tianjin University

135 Yaguan Road, Jinnan District

Tianjin, 300354

China

Phone: 86 15034071215

Email: 2020212056@tju.edu.cn

Abstract

Background: Mental health problems are a crucial global public health concern. Owing to their cost-effectiveness and accessibility, conversational agent interventions (CAIs) are promising in the field of mental health care.

Objective: This study aims to present a thorough summary of the traits of CAIs available for a range of mental health problems, find evidence of efficacy, and analyze the statistically significant moderators of efficacy via a meta-analysis of randomized controlled trial.

Methods: Web-based databases (Embase, MEDLINE, PsycINFO, CINAHL, Web of Science, and Cochrane) were systematically searched dated from the establishment of the database to October 30, 2021, and updated to May 1, 2022. Randomized controlled trials comparing CAIs with any other type of control condition in improving depressive symptoms, generalized anxiety symptoms, specific anxiety symptoms, quality of life or well-being, general distress, stress, mental disorder symptoms, psychosomatic disease symptoms, and positive and negative affect were considered eligible. This study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Data were extracted by 2 independent reviewers, checked by a third reviewer, and pooled using both random effect models and fixed effects models. Hedges g was chosen as the effect size.

Results: Of the 6900 identified records, a total of 32 studies were included, involving 6089 participants. CAIs showed statistically significant short-term effects compared with control conditions in improving depressive symptoms ($g=0.29$, 95% CI 0.20-0.38), generalized anxiety symptoms ($g=0.29$, 95% CI 0.21-0.36), specific anxiety symptoms ($g=0.47$, 95% CI 0.07-0.86), quality of life or well-being ($g=0.27$, 95% CI 0.16-0.39), general distress ($g=0.33$, 95% CI 0.20-0.45), stress ($g=0.24$, 95% CI 0.08-0.41), mental disorder symptoms ($g=0.36$, 95% CI 0.17-0.54), psychosomatic disease symptoms ($g=0.62$, 95% CI 0.14-1.11), and negative affect ($g=0.28$, 95% CI 0.05-0.51). However, the long-term effects of CAIs for the most mental health outcomes were not statistically significant ($g=-0.04$ to 0.39). Personalization and empathic response were 2 critical facilitators of efficacy. The longer duration of interaction with conversational agents was associated with the larger pooled effect sizes.

Conclusions: The findings show that CAIs are research-proven interventions that ought to be implemented more widely in mental health care. CAIs are effective and easily acceptable for those with mental health problems. The clinical application of this novel digital technology will conserve human health resources and optimize the allocation of mental health services.

Trial Registration: PROSPERO CRD42022350130; <https://tinyurl.com/mvkh6w9p>

(*J Med Internet Res* 2023;25:e43862) doi: [10.2196/43862](https://doi.org/10.2196/43862)

KEYWORDS

chatbot and conversational agent; mental health; meta-analysis; depression; anxiety; quality of life; stress; mobile health; mHealth; digital medicine; meta-regression; mobile phone

Introduction

Background

To promote mental health for everyone, everywhere, the World Health Organization's most recent global mental health report strives to motivate and guide revolutionary action [1,2]. According to estimates by the World Health Organization, in the first year of the COVID-19 pandemic, the incidence of both depression and anxiety disorders increased by more than 25% [3]. However, mental health services are few and still underutilized globally [4-6]. This was because traditional face-to-face mental health care still had many limitations, such as expensive treatment, lack of experienced therapists, poor service quality, geographical constraints, and delayed treatment [7,8], and the resulting stigma and discrimination are also considered the most important barriers to providing mental health services [9,10].

Because of their increased acceptance and accessibility, digital mental health interventions have emerged as an important research area with evidence-based psychotherapies implemented on digital platforms [11]. Conversational agent interventions (CAIs) [12] were the new wave of digital mental health interventions to cope with the insufficient and inadequate mental health services [1,13]. Substantial human health resources will be saved if the CAIs are proven effective and suitable for widespread implementation [14].

Software programs known as conversational agents (CAs) use artificial intelligence techniques to simulate human behavior and offer a task-oriented framework with evolving dialogue able to engage in conversation [15]. CAs are equipped with computer models that range from succinct decision trees, where different responses to a questionnaire result in different responses from the CAs to machine learning-based and natural language processing-based algorithms that classify real-time multimodal input into a user's emotional state, allowing the CAs to respond empathically [16,17]. One of the most crucial features of CAs in mental health was interactivity, which was designed to promote a conversational process instead of a single psychological education, in which inputs and outputs are generated in unrestricted natural language rather than predefined or preprogrammed choices or messages [12]. The other was automation, which means that most CAs for mental health issues can independently provide automated services to users without the participation and guidance of human [18].

Certain studies claimed that CAIs can help users feel accompanied and understood [19,20]; several studies found that users establish therapeutic bonds with chatbots [21-24]; and other studies revealed some hazards associated with CAIs, including *misunderstanding* that may result in inefficient or even harmful interventions, a lack of crisis warning systems, and a lack of privacy protection [25]. An increasing number of CAIs have emerged in recent years, as the digital medical and mobile health fields have grown [26], solving and improving a

larger range of mental health issues in addition to depression and anxiety. Therefore, a thorough systematic review and meta-analysis of CAIs for mental health problems are urgently needed.

Objectives

In this study, we outlined the clinical and nonclinical features of CAIs in mental health using a systematic review. We then evaluated the short- and long-term efficacy of CAIs for different mental health outcomes via a meta-analysis and assessed whether various characteristics related to the intervention and sample moderated the observed effect sizes.

Methods

Study Protocol and Registration

We have reported the findings in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Table S1 in [Multimedia Appendix 1](#) [27-58]) guidelines [59]. The study protocol was registered in the PROSPERO database (CRD42022350130).

Search Strategy

As mentioned in the initial registration, we searched 6 major web-based databases (Embase, MEDLINE, PsycINFO, CINAHL, Web of Science, and Cochrane), dated from the establishment of the database to October 30, 2021, and updated to May 1, 2022, using the following search terms: (“conversational agent*” OR “conversational system*” OR “dialog system*” OR “dialogue system*” OR “assistance technolog*” OR “relational agent*” OR chatbot* OR “automat*” OR “virtual human*” OR “virtual agent” OR “virtual coach” OR “virtual therap*” OR avatar OR “artificial Intelligence”) AND (“depress*” OR “anxiety” OR “agoraphobia” OR “phobia*” OR “panic” OR “mental health” OR “mental illness*” OR “mental disorder” OR psycholog* OR “affective disorder*” OR “bipolar” OR “mood disorder*” OR “psychosis” OR “psychotic” OR “schizophre*” OR “well-being” OR “well-being” OR “quality of life” OR “self-harm” or “self-injury” OR “stress*” OR “distress*” OR “mood” OR “loneliness” OR “social isolation” OR “autism” OR “suicid*” OR “cogniti*” OR insomnia OR emotion* OR affect*) AND (“randomized trial*” OR “controlled trial*” OR “randomised trial*” OR RCT OR RCTs). In addition, the reference lists of the included original studies and previous reviews were manually searched to identify any further eligible studies. To reduce the possibility of publication bias, eligible published and unpublished papers were searched for inclusion.

Eligibility Criteria

Participants

Participants were categorized as (1) clinical sample (from outpatients or inpatients in psychiatric hospitals), (2) symptomatic sample (from those at a diagnostic or subthreshold level of mental health problems), and (3) general sample (from

universities, companies, or communities with a wider range of recruitment scales and standards). People with physical illnesses who did not have mental health problems were excluded.

Interventions

Interventions with four types of CAs were eligible: (1) chatbot, a software program that simulates conversations with users through text or voice depending on artificial intelligence [60], such as ELIZA (Massachusetts Institute of Technology), the first chatbot, via which users can input text to simulate a conversation with a Rogerian psychotherapist [61]; (2) embodied CA (ECA), the digital characters that simulate key properties of human face-to-face conversation, both verbal and nonverbal (eg, speech, gestures, and facial expressions) [62], called virtual human [63] and digital human [64]; (3) CA in virtual reality (VR), which offers a safe, convenient, and accessible medium to the controlled exposure to anxiety-inducing stimuli within a VR environment to deliver exposure-based behavioral treatments [65]; and (4) avatar, referring to the CA in avatar therapy, a new wave of relational approaches, now used mainly to treat auditory verbal hallucinations and depressive symptoms, allowing patients to interact with a digital representation (avatar) whose speech closely matches the pitch and tone of the persecutory voice and gradually gains increased power and control within the relationship [27].

Comparators

Control conditions were categorized as (1) active controls (therapist-led interventions, other CAIs, or treatment as usual), (2) information or attentional controls (self-help e-book or text), and (3) passive controls (waitlist [WL] or assessment only).

Outcomes

Eligible studies were those that reported at least 1 mental health outcome and provided the outcome data required to calculate the effect size. Studies were excluded if they focused on physical illness or physical health. We selected and analyzed the following mental health outcomes: depressive symptoms, generalized anxiety symptoms, specific anxiety symptoms (phobia symptoms, social anxiety symptoms, and panic symptoms), quality of life or well-being, general distress, stress, mental disorder symptoms (substance use disorder symptoms, attention-deficit/hyperactivity disorder symptoms, and psychotic symptoms), psychosomatic disease symptoms (chronic pain symptoms and irritable bowel syndrome symptoms), and positive and negative affect.

Study Design

The eligible studies were randomized controlled trials (RCTs). Studies were excluded if they met the following criteria: (1) aimed at assessment, management, medical skill training, or health knowledge access and (2) only focused on technology, engagement, usability, or user experience.

Data Extracting

The following data were extracted: (1) basic study characteristics (including the first author, publication date, country, and mental health problems), (2) participant characteristics (including target sample, sample size, age, and gender), (3) intervention

characteristics, (4) outcome measures, and (5) engagement and user experience.

Study results related to mental health outcomes were extracted in the form of means and SDs at baseline, postintervention, and follow-up when available, or, if not available, effect sizes. The study authors were contacted in case of missing or unclear information, and the study was excluded if they failed to provide the data.

It should be noted that in digital interventions, *interaction frequency* was usually used to represent dosage [66]. However, because different studies had different operational definitions of interaction frequency, to preliminarily explore the dose-response relationships of CAIs in mental health, we uniformly used the comparable data, average duration of interaction per day, to quantify the dosage in meta-regression.

Quality Assessment

The quality of the trials was independently evaluated by 2 authors using the Cochrane Collaboration tool [67] to assess the risk of bias. The Cochrane Collaboration tool evaluates 7 domains (sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective outcome reporting, and other risk of bias). For each domain, a rating of low (+), high (−), or unclear (?) was made for each trial. Only trials that received a low-risk rating on all 7 criteria were considered to have a low risk of bias. Trials were considered to have a high risk of bias if they were rated high in any bias domain other than performance bias, as blinding of participants and personnel is almost impossible in CAIs studies [68]. The risk of bias graph was drawn using Review Manager (version 5.4; The Cochrane Collaboration).

Meta-analysis

Given the differences at baseline, the change-from-baseline scores were computed for each group to represent the pre-post efficacy. For each comparison between the treatment group and control group, the standardized mean difference was calculated, as there were different scales from the included study to measure the same outcome.

We chose Hedges g as the effect sizes [69]. A positive g indicates that the CAIs condition had better outcomes than the comparison condition. If means and SDs were not reported, the effect sizes were converted via other available statistics (eg, d or η^2). If data from both intention-to-treat and completer analyses were presented, the former were extracted and analyzed. We analyzed the pooled effect sizes of outcomes at postintervention as the short-term efficacy and the pooled effect sizes of outcomes at follow-up as the long-term efficacy.

Stata (version 15; Stata Corp) was used to perform the analyses. Heterogeneity between studies was quantified by calculating the I^2 statistic and Cochran Q statistic [70]. A random effects model was applied when substantial heterogeneity was observed ($P < .05$ or $I^2 > 50\%$); otherwise, a fixed effects model was used [71].

In a few studies, the same intervention condition was compared with >1 control condition (or vice versa), which may have artificially reduced the heterogeneity estimate and affected the pooled effect size, as these comparisons were not independent of each other. Thus, sensitivity analyses were run: only the comparison with the smallest effect size included in the meta-analysis or the largest effect size, which ensured that only 1 effect size per study was included in the analysis. We also performed an analysis only including low-risk articles to determine the pooled effect size after controlling for the risk of bias.

We conducted a subgroup analysis and meta-regression with the primary outcomes (depressive symptoms, generalized anxiety symptoms, specific anxiety symptoms, quality of life, general distress, and stress). Subgroup analyses were performed using Comprehensive Meta-Analysis (version 3; Bio-stat Inc) under a mixed effects model, which pools studies within a subgroup using a random effects model but tests for statistically

significant differences between subgroups using fixed effects models [72]. As avatar does not fit the definition of the self-guided intervention, we pooled the effect sizes of the chatbot, ECA, and CA in VR to show the effect of self-guided CAIs in subgroup analysis.

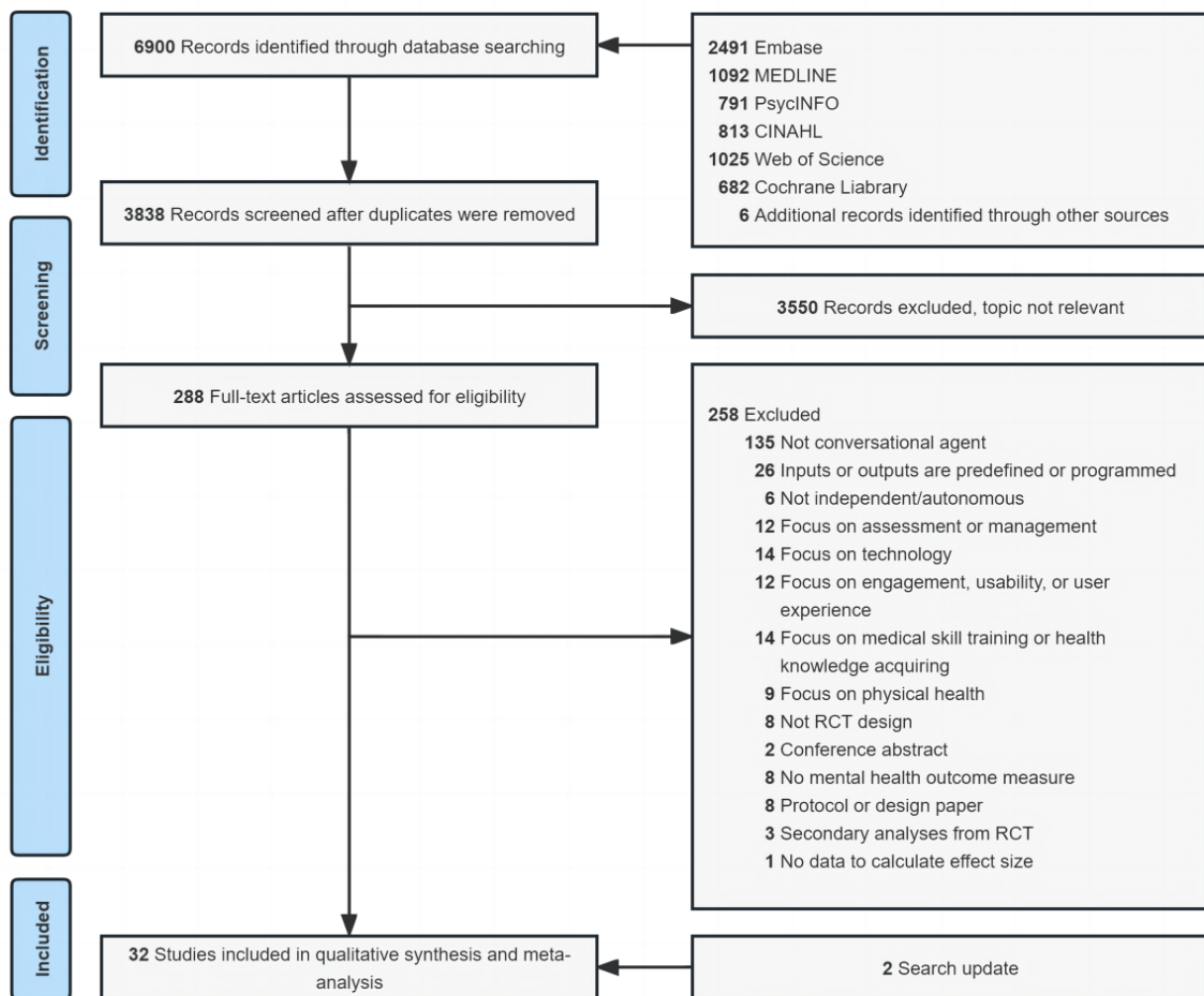
Univariate random effects meta-regression used residual restricted maximum likelihood to measure between-study variance (τ^2) with a Knapp-Hartung modification as recommended models [73].

We applied different methods to examine publication bias (funnel plot, Begg and Mazumdar rank correlation test [74], Egger regression test [75], and Duval and Tweedie trim-and-fill procedure [76]).

Results

The details of the selection process are presented in a PRISMA flow diagram (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) study flow diagram. RCT: randomized controlled trial.



Study Characteristics

The 32 RCTs with 26 CAs were included, involving 6089 participants, with sample sizes ranging from 19 to 2668 (Table

S2 in Multimedia Appendix 1). The average age was 36.32 (SD 13.44; range 21.40-71.47) years, and 75.41% (2922/3875) of the participants were women from the reported data. A total of 16 studies with CAs were classified as chatbot [28-43], 6 studies

with CAs classified as ECA [44-49], 5 studies with CAs classified as CA in VR [50-54], and 5 studies with CAs classified as avatar [27,55-58].

Among the included studies, 9 were conducted in the United States; 6 were conducted in the United Kingdom; 3 were conducted in Sweden; 2 each conducted in Japan, Korea, Switzerland, and Canada; and 1 each conducted in Argentina, China, Ireland, the Netherlands, Germany, and New Zealand. A total of 8 studies recruited clinical samples, 14 recruited symptomatic samples, and 10 recruited general samples. In addition, 17 studies were based on cognitive behavioral therapy (CBT), and 15 studies were based on other theories such as patient-centered therapy, acceptance and commitment therapy, method of levels therapy, and problem-solving treatment. A total of 18 studies supported personalization and tailoring, and 14 studies did not support personalization and tailoring. Moreover, 18 studies supported emotional and empathic responses, and 14 studies did not support emotional and empathic responses. In total, 15 studies provided automatic reminders to engage, and 17 studies did not provide automatic reminders to engage. Nine studies directly aimed at depressive symptoms, 11 studies directly aimed at generalized anxiety symptoms, 6 studies directly aimed at specific anxiety symptoms, 2 studies directly aimed at quality of life, 5 studies directly aimed at general distress, 2 studies directly aimed at stress, and 4 studies directly aimed at other outcomes. Eleven studies had an intervention length of 0 to 4 weeks, 10 studies had an intervention length of 5 to 8 weeks, and 11 studies had an intervention length of >9 weeks. Seventeen studies had no follow-up, 7 studies had a follow-up length of 0 to 8 weeks, and 8 studies had a follow-up length of ≥9 weeks. Fourteen studies used passive controls, 8 studies used information or attentional controls, and 10 studies used active controls.

Attrition rates between baseline and postintervention measures were reported in 31 studies and varied widely from no attrition to 82.98% (2214/2668) of participants. More dropouts were from the control condition (2081/3583, 58.08%) than from the intervention condition (740/1888, 39.19%). The ration of the interaction reported in 16 studies ranged from 0.57 to 6.43 (mean 4.48, SD 2.11) minutes per day. Most studies (14/17, 82%) have reported good acceptability and usability of CAIs. Fifteen studies reported user experience, of which 8 performed thematic analysis on the feedback of participants.

Risk of Bias

Interrater reliability suggested substantial agreement between the raters for 7 domains of the Cochrane Collaboration tool (Cohen κ =0.86, 0.79, 0.67, 0.89, 0.93, 0.75, and 0.83, respectively). Eight studies were assessed as having a low risk of bias, 11 studies had some risk of bias, and 13 studies had a high risk of bias (Figures S1 and S2 in [Multimedia Appendix 1](#)).

The funnel plot of short-term effects (Figure S3 in [Multimedia Appendix 1](#)) and long-term effects (Figure S4 in [Multimedia Appendix 1](#)) and the results of Begg and Egger tests performed well on most mental health outcomes (Table S3 in [Multimedia Appendix 1](#)).

Efficacy

Depressive Symptoms

The pooled effect size for the 27 postintervention comparisons between CAIs and control conditions on depressive symptoms was $g=0.29$ (95% CI 0.20-0.38), with moderate heterogeneity ($I^2=42.90\%$, 95% CI 9.79%-63.86%; [Figures 2 and 3](#); [Table 1](#)). This effect size was slightly smaller after adjusting for any potential publication bias ($g=0.26$) and slightly larger when restricting the analyses to trials with a low risk of bias ($g=0.32$; [Table 1](#)). The pooled effect size of long-term efficacy for the 22 follow-up comparisons between CAIs and control conditions on depressive symptoms was $g=0.16$ (95% CI 0.06-0.26), with low heterogeneity ($I^2=8.05\%$, 95% CI 0.00%-42.12%; [Figure 2](#); [Table 1](#); [Figure S5 in Multimedia Appendix 1](#)).

The subgroup analyses revealed 5 statistically significant moderators ([Figure 4](#); [Table 1](#)). Studies that directly aimed at depressive symptoms produced larger effect sizes than those that did not ($P=.004$). Studies with a follow-up length between 0 and 8 weeks or between 9 and 16 weeks produced larger effect sizes than those with a follow-up length of ≥17 weeks ($P=.049$). Studies that supported personalization and tailoring produced larger effect sizes than those that did not ($P=.045$). Studies that supported emotional and empathic responses produced larger effect sizes than those that did not ($P=.008$). Studies that provided automatic reminders to engage produced smaller effect sizes than those that did not ($P=.04$).

Meta-regression analyses revealed a statistically significant effect of dosage ($b=0.160$, 95% CI 0.014-0.306; $P=.04$) on the pooled effect size ([Table 2](#); [Figure S6 in Multimedia Appendix 1](#)).

Figure 2. Overall effect analyses with short- and long-term efficacy of CAIs for mental health problems. Hedges g scores (mean and 95% CI) are given (positive values indicate better performance among individuals at CAIs vs control individuals), along with the number of comparisons (n) included and sample size. CAI: conversational agent intervention.

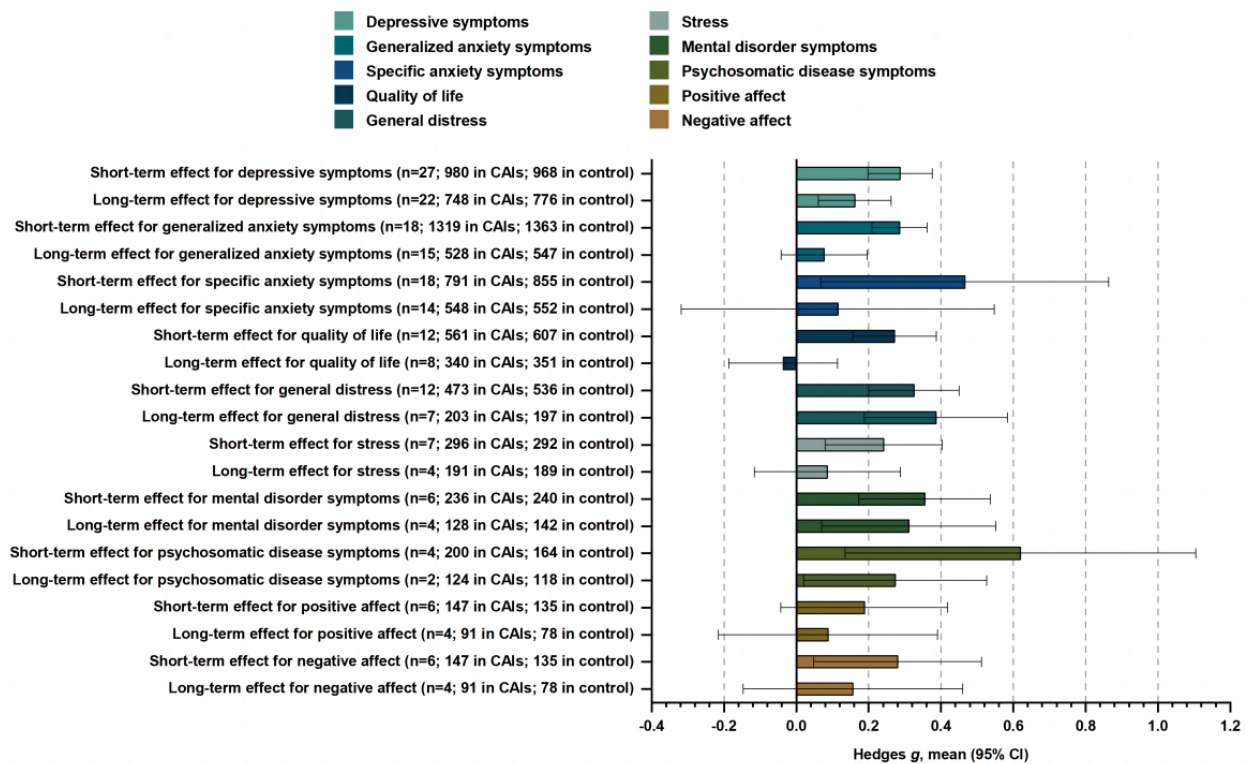


Figure 3. Forest plot for the short-term effects of conversational agent interventions on depressive symptoms.

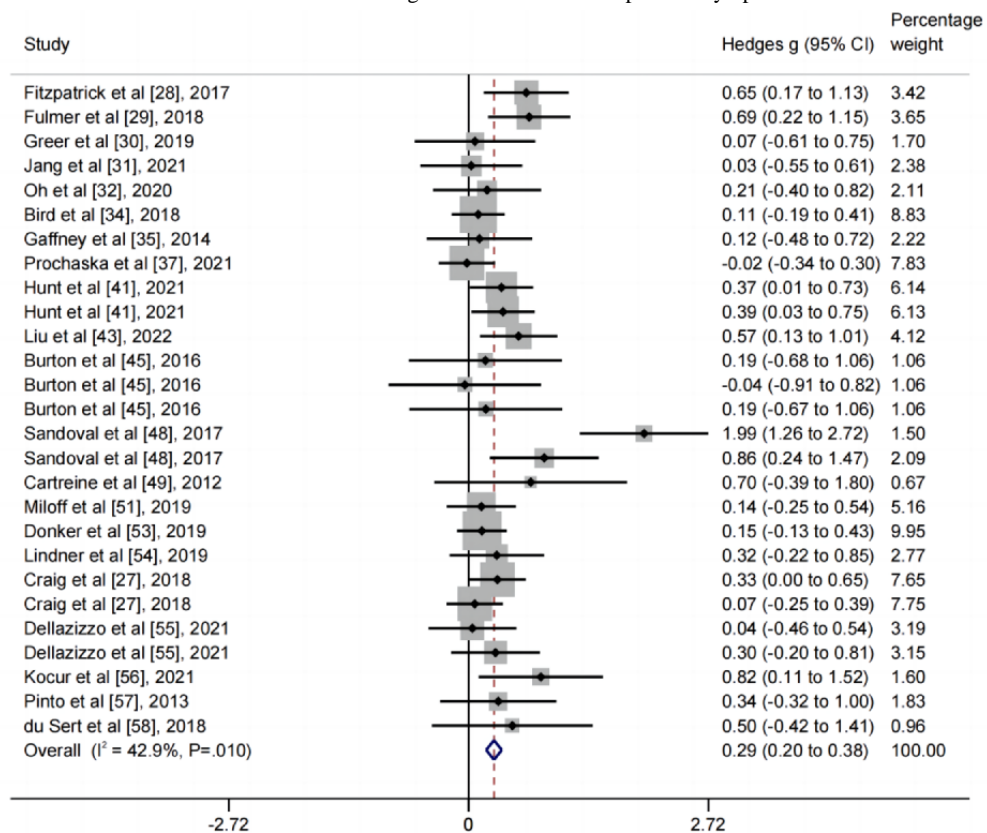


Table 1. Meta-analysis of efficacy of conversational agent interventions for depressive and generalized anxiety symptoms.

	Depressive symptoms				Generalized anxiety symptoms			
	Value, n ^a	g (95% CI)	I ² (%; 95% CI)	P value ^b	Value, n ^a	g (95% CI)	I ² (%; 95% CI)	P value ^b
Overall effect analysis				N/A ^c				N/A
Short-term effect	27	0.29 (0.20 to 0.38) ^d	42.90 (9.79 to 63.86)		18	0.29 (0.21 to 0.36) ^d	27.32 (0.00 to 58.94)	
Adjusted for publication bias	30	0.26 (0.11 to 0.41) ^e	N/A		18	0.29 (0.21 to 0.36) ^d	N/A	
Long-term effect	22	0.16 (0.06 to 0.26) ^e	8.05 (0.00 to 42.12)		15	0.08 (-0.04 to 0.20)	0.00 (0.00 to 53.61)	
Adjusted for publication bias	23	0.15 (0.05 to 0.25) ^e	N/A		19	0.00 (-0.11 to 0.11)	N/A	
Sensitivity analysis				N/A				N/A
One effect size per study (largest)	21	0.36 (0.21 to 0.50) ^d	50.44 (18.09 to 70.01)		17	0.28 (0.19 to 0.36) ^d	30.76 (0.00 to 61.45)	
One effect size per study (smallest)	21	0.26 (0.17 to 0.36) ^d	22.48 (0.00 to 54.46)		17	0.24 (0.15 to 0.32) ^d	10.92 (0.00 to 47.69)	
Low risk of bias only (all criteria met)	7	0.32 (0.18 to 0.46) ^d	6.07 (0.00 to 41.68)		5	0.18 (0.01 to 0.35) ^f	0.00 (0.00 to 79.20)	
Subgroup analyses								
Conversational agent type				.48				.46
Chatbot	11	0.29 (0.14 to 0.44) ^d	25.76 (0.00 to 63.21)		13	0.29 (0.17 to 0.41) ^d	31.89 (0.00 to 64.80)	
Embodied conversational agent	6	0.68 (0.05 to 1.30) ^f	71.65 (34.28 to 87.77)		0	— ^g	—	
Conversational agent in virtual reality	3	0.18 (-0.04 to 0.39)	0.00 (0.00 to 89.60)		3	0.23 (0.02 to 0.46) ^f	11.24 (0.00 to 51.16)	
Avatar	7	0.25 (0.08 to 0.42) ^e	0.00 (0.00 to 70.81)		2	0.10 (-0.17 to 0.37)	0.00	
Self-guided				.41				.22
Yes	20	0.35 (0.19 to 0.51) ^d	52.85 (0.22 to 0.72)		16	0.28 (0.18 to 0.38) ^d	26.74 (0.00 to 59.85)	
No	7	0.25 (0.08 to 0.42) ^e	0.00 (0.00 to 70.81)		2	0.10 (-0.17 to 0.37)	0.00	
Control condition type				.05				.06
Waitlist or assessment only	8	0.48 (0.14 to 0.82) ^e	76.67 (53.55 to 88.29)		6	0.35 (0.25 to 0.44) ^d	0.00 (0.00 to 74.63)	
Information or attentional control	6	0.47 (0.26 to 0.68) ^d	0.00 (0.00 to 74.63)		6	0.27 (0.01 to 0.54) ^f	38.56 (0.00 to 75.59)	
Active control	13	0.20 (0.07 to 0.33) ^e	0.00 (0.00 to 56.60)		6	0.12 (-0.05 to 0.28)	0.00 (0.00 to 74.63)	
Intervention target				.004 ^h				.03
Directly aimed at this outcome	12	0.61 (0.34 to 0.88) ^d	49.82 (2.62 to 74.14)		12	0.33 (0.22 to 0.43) ^d	18.95 (0.00 to 57.96)	
Not directly aimed at this outcome	15	0.18 (0.08 to 0.29) ^d	0.00 (0.00 to 53.61)		6	0.11 (-0.05 to 0.28)	0.00 (0.00 to 74.63)	
Intervention length (weeks)				.30				.44

	Depressive symptoms				Generalized anxiety symptoms			
	Value, n ^a	<i>g</i> (95% CI)	<i>I</i> ² (%; 95% CI)	<i>P</i> value ^b	Value, n ^a	<i>g</i> (95% CI)	<i>I</i> ² (%; 95% CI)	<i>P</i> value ^b
0-4	8	0.57 (0.18 to 0.96) ^e	75.12 (49.90 to 87.64)		4	0.15 (-0.10 to 0.39)	25.32 (0.00 to 71.30)	
5-8	10	0.28 (0.11 to 0.45) ^e	0.00 (0.00 to 62.37)		6	0.25 (0.07 to 0.49) ^f	33.15 (0.00 to 73.08)	
≥9	9	0.25 (0.11 to 0.38) ^d	2.12 (0.00 to 23.51)		8	0.32 (0.21 to 0.43) ^d	15.57 (0.00 to 58.54)	
Follow-up lengthⁱ (weeks)				<i>.049</i>				<i>.74</i>
0-8	5	0.30 (-0.01 to 0.61)	29.97 (0.00 to 72.95)		4	0.05 (-0.21 to 0.32)	0.00 (0.00 to 84.69)	
9-16	8	0.26 (0.10 to 0.41) ^e	0.00 (0.00 to 67.58)		5	0.15 (-0.03 to 0.34)	0.00 (0.00 to 79.20)	
≥17	9	0.01 (-0.15 to 0.16)	0.00 (0.00 to 64.80)		6	0.05 (-0.18 to 0.27)	23.24 (0.00 to 67.22)	
Target sample				<i>.21</i>				<i>.005</i>
Clinical sample	10	0.20 (0.04 to 0.36) ^f	0.00 (0.00 to 62.37)		4	0.04 (-0.18 to 0.27)	0.00 (0.00 to 84.69)	
Symptomatic sample	13	0.42 (0.20 to 0.63) ^d	66.43 (39.76 to 81.29)		10	0.21 (0.09 to 0.34) ^e	2.65 (0.00 to 27.71)	
Nonclinical or nonsymp-tomatic sample	4	0.42 (0.13 to 0.70) ^e	0.00 (0.00 to 84.69)		4	0.40 (0.29 to 0.51) ^d	0.00 (0.00 to 84.69)	
Personalization and tailoring				<i>.05</i>				<i>.04</i>
Yes	17	0.44 (0.24 to 0.64) ^d	59.96 (30.18 to 76.04)		12	0.33 (0.21 to 0.44) ^d	21.36 (0.00 to 59.62)	
No	10	0.17 (0.04 to 0.30) ^e	0.00 (0.00 to 62.37)		6	0.13 (-0.02 to 0.28)	0.00 (0.00 to 74.63)	
Emotional and empathic responses				<i>.008</i>				<i>.02</i>
Yes	16	0.46 (0.27 to 0.66) ^d	60.63 (32.01 to 77.20)		10	0.34 (0.23 to 0.44) ^d	18.19 (0.00 to 59.10)	
No	11	0.14 (-0.04 to 0.28)	0.00 (0.00 to 60.23)		8	0.12 (-0.03 to 0.27)	0.00 (0.00 to 67.58)	
Automatic reminders to engage provided				<i>.04</i>				<i>.22</i>
Yes	11	0.19 (0.05 to 0.33) ^e	5.89 (0.00 to 41.41)		10	0.20 (0.04 to 0.35) ^f	21.20 (0.00 to 61.32)	
No	16	0.43 (0.24 to 0.61) ^d	51.26 (13.59 to 72.51)		8	0.32 (0.20 to 0.43) ^d	20.00 (0.00 to 62.47)	
Cognitive behavioral therapy-based conversational agent				<i>.87</i>				<i>.93</i>
Yes	14	0.32 (0.18 to 0.46) ^d	15.14 (0.00 to 53.48)		9	0.25 (0.10 to 0.40) ^e	13.82 (0.00 to 55.85)	

	Depressive symptoms				Generalized anxiety symptoms			
	Value, n ^a	<i>g</i> (95% CI)	<i>I</i> ² (%; 95% CI)	<i>P</i> value ^b	Value, n ^a	<i>g</i> (95% CI)	<i>I</i> ² (%; 95% CI)	<i>P</i> value ^b
No	13	0.34 (0.12 to 0.56) ^e	60.05 (26.57 to 78.26)		9	0.26 (0.12 to 0.40) ^d	41.55 (0.00 to 73.06)	

^aNumber of comparisons.

^b*P* value represents the significance of the *Q* test.

^cN/A: not applicable.

^d*P* < .001.

^e*P* < .01.

^f*P* < .05.

^gMissing data.

^hItalicized values indicate statistically significant differences.

ⁱFollow-up length is a moderator of long-term effects, whereas the other variables are moderators of short-term effects.

Figure 4. Subgroup analyses of efficacy of CAIs for depressive symptoms. *P* value represents the significance of the *Q* test. CA: conversational agent; CAI: conversational agent intervention; CBT: cognitive behavioral therapy; ECA: embodied conversational agent; VR: virtual reality; WL: waitlist.

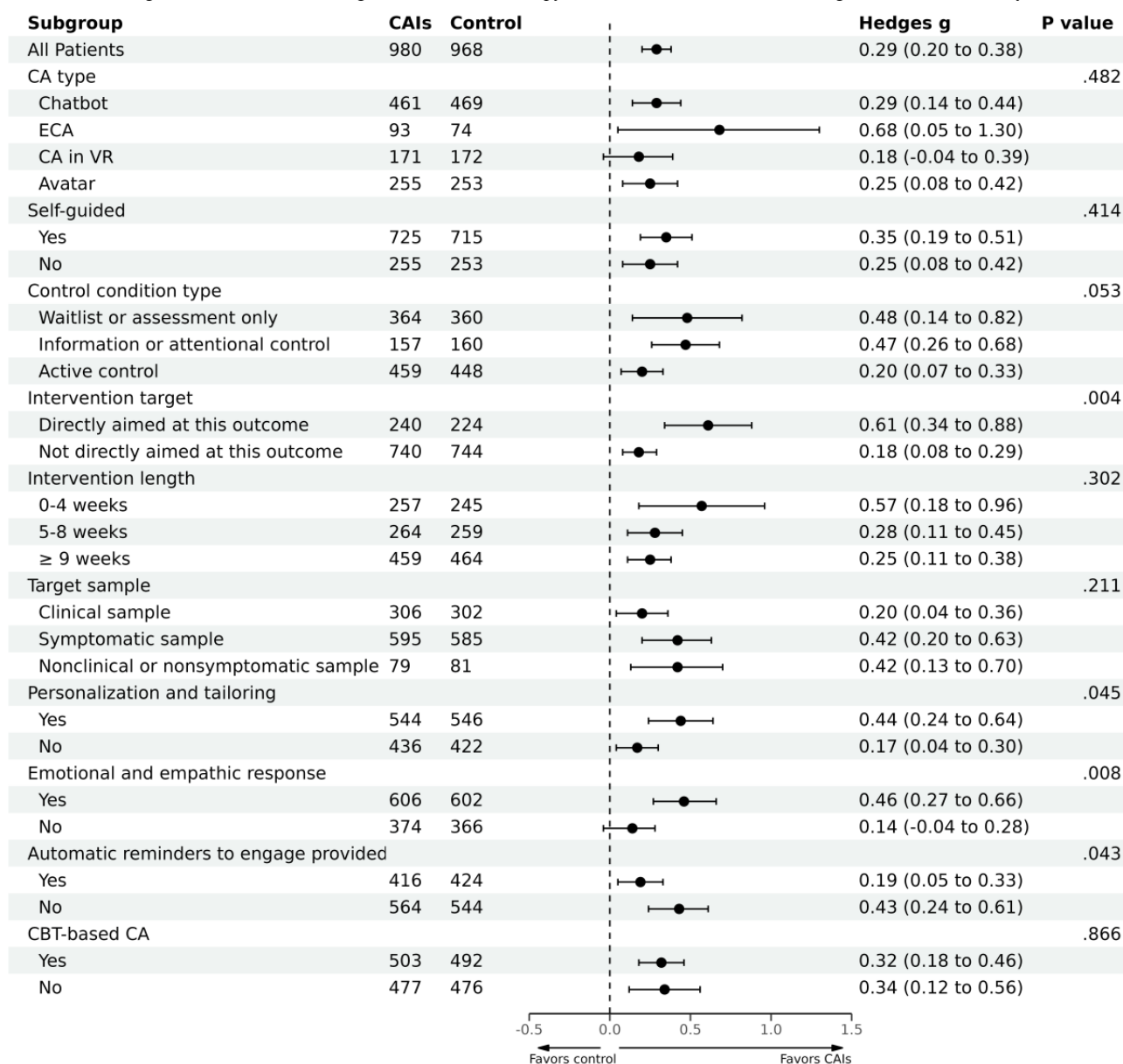


Table 2. Meta-regression of efficacy of conversational agent interventions for primary mental health outcomes.

Dependent and covariate	Coefficient (95% CI)	SE	<i>P</i> value
Depressive symptoms			
Age	-0.011 (-0.027 to 0.005)	0.008	.18
Gender	0.080 (-0.712 to 0.872)	0.385	.84
Dose ^a	0.160 (0.014 to 0.306)	0.067	.04 ^b
Generalized anxiety symptoms			
Age	-0.007 (-0.022 to 0.008)	0.007	.33
Gender	0.438 (0.029 to 0.847)	0.193	.04
Dose ^a	0.067 (0.006 to 0.127)	0.025	.04
Specific anxiety symptoms			
Age	0.084 (0.014 to 0.154)	0.033	.01
Gender	-3.414 (-6.118 to -0.710)	1.276	.02
Dose ^a	0.186 (-0.106 to 0.474)	0.122	.10
General distress			
Age	-0.003 (-0.014 to 0.008)	0.005	.50
Gender	-1.078 (-4.097 to 1.940)	1.355	.45
Dose ^a	0.021 (-0.744 to 0.786)	0.178	.92
Quality of life			
Age	0.008 (-0.017 to 0.032)	0.011	.52
Gender	0.151 (-0.975 to 1.277)	0.505	.71
Dose ^a	0.102 (-0.429 to 0.633)	0.167	.54
Stress			
Age	0.014 (-0.007 to 0.035)	0.008	.16
Gender	0.251 (-1.498 to 1.999)	0.680	.73
Dose ^a	-0.207 (-16.303 to 15.890)	1.267	.90

^aAverage duration of interaction with conversational agent per day.

^bItalicized values indicate statistically significant differences.

Generalized Anxiety Symptoms

The pooled effect size for the 18 postintervention comparisons was $g=0.29$ (95% CI 0.21-0.36), with moderate heterogeneity ($I^2=27.32\%$, 95% CI 0.00%-58.94%; [Figures 2 and 5](#); [Table 1](#)). It remained significant across all sensitivity analyses ([Table 1](#)). The pooled effect size of long-term efficacy was $g=0.08$ (95% CI -0.04 to 0.20), with low heterogeneity ($I^2=0.00\%$, 95% CI

0.00%-53.61%; [Figure 2](#); [Table 1](#); [Figure S7 in Multimedia Appendix 1](#)).

The subgroup analyses revealed 4 statistically significant moderators ([Figure 6](#); [Table 1](#)). Larger effect sizes were found in studies that directly aimed at generalized anxiety symptoms ($P=.03$), that with symptomatic sample or general sample ($P=.005$), that supported personalization and tailoring ($P=.04$), and that supported emotional and empathic responses ($P=.02$).

Figure 5. Forest plot for the short-term effects of conversational agent interventions on generalized anxiety symptoms.

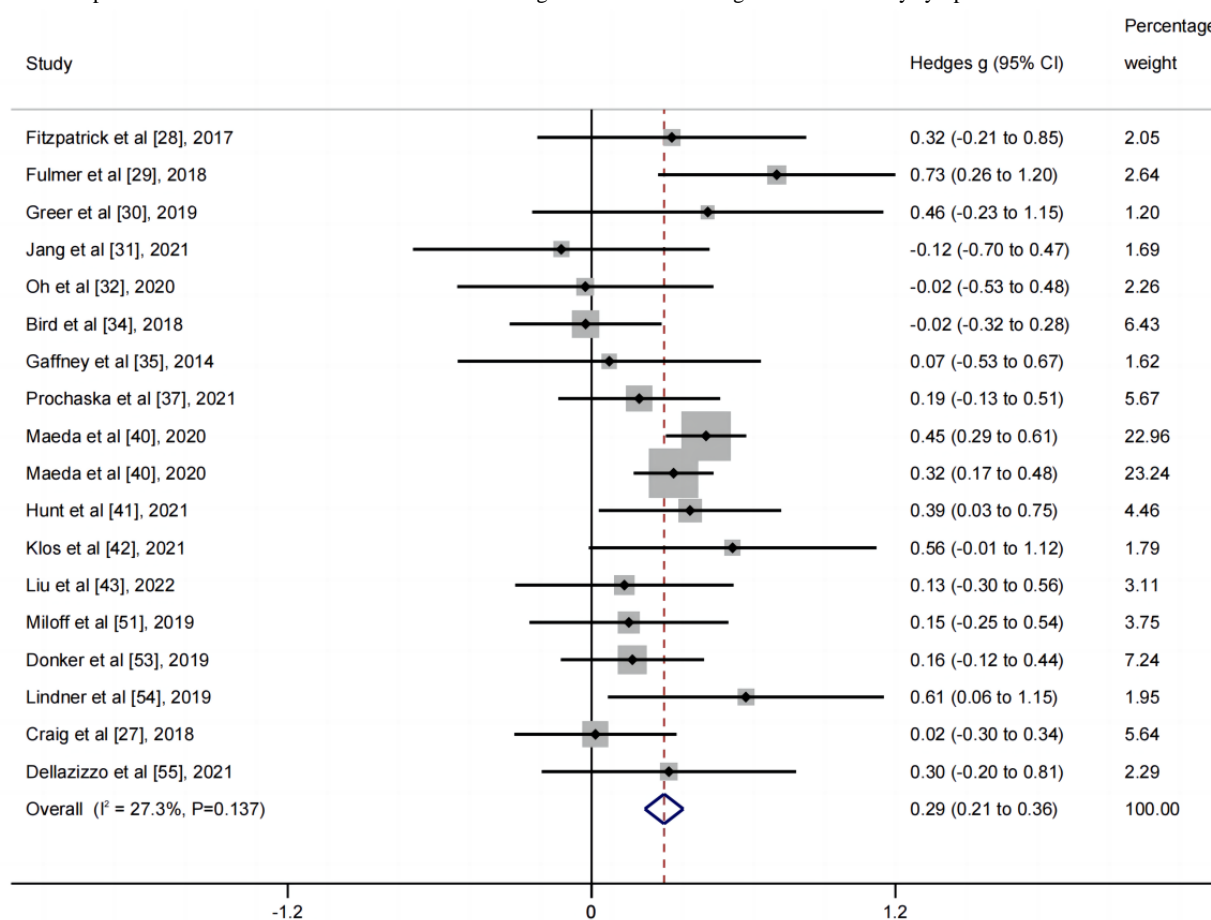
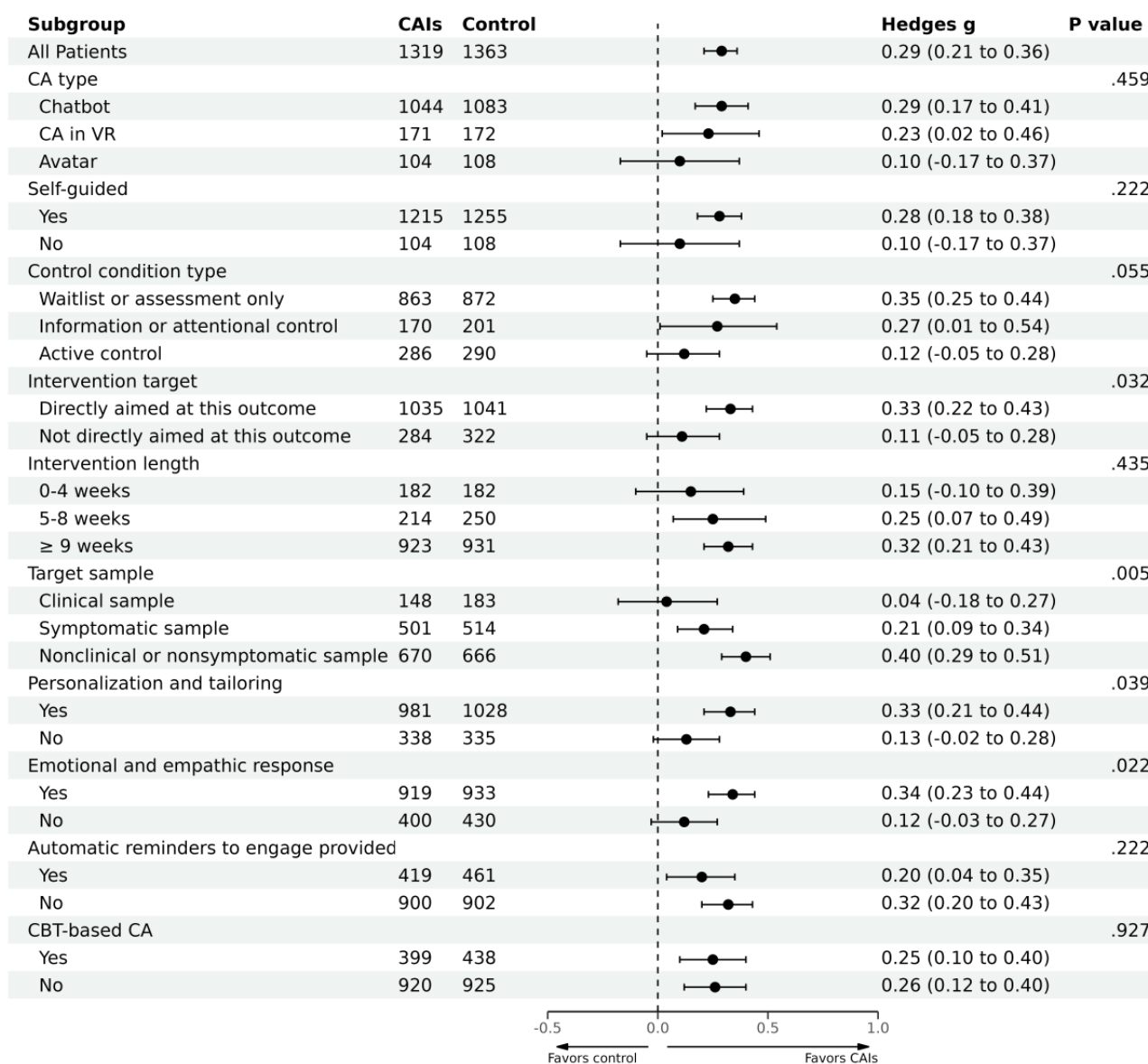


Figure 6. Subgroup analyses of efficacy of CAIs for generalized anxiety symptoms. *P* value represents the significance of the Q test. CA: conversational agent; CAI: conversational agent intervention; CBT: cognitive behavioral therapy; ECA: embodied conversational agent; VR: virtual reality; WL: waitlist.



Meta-regression analyses revealed statistically significant effects of gender ($b=0.448$, 95% CI 0.024-0.873; $P=.04$) and dosage ($b=0.067$, 95% CI 0.006-0.127; $P=.04$) on the pooled effect size (Table 2; Figure S6 in Multimedia Appendix 1).

Specific Anxiety Symptoms

The pooled effect size for the 18 postintervention comparisons was $g=0.47$ (95% CI 0.07-0.86), with high heterogeneity ($I^2=93.17\%$, 95% CI 90.62%-95.03%; Figure 2; Table 3; Figure S8 in Multimedia Appendix 1). It did not remain significant when compared with the smallest effect size and when restricting the analyses to trials with a low risk of bias (Table 3). The pooled effect size of long-term efficacy was $g=0.11$ (95% CI -0.32 to 0.55), with high heterogeneity ($I^2=91.82\%$, 95% CI

88.01%-94.42%; Figure 2; Table 3; Figure S9 in Multimedia Appendix 1).

Subgroup analyses revealed 5 statistically significant moderators (Table 3; Figure S10 in Multimedia Appendix 1). Larger effect sizes were found in studies with WL or assessment controls ($P<.001$), longer intervention lengths (5-8 weeks; $P<.001$), shorter follow-up length (0-8 weeks; $P<.001$), CBT-based CAs ($P=.008$), and studies that supported emotional and empathic responses ($P<.001$).

Meta-regression analyses revealed statistically significant effects of age ($b=0.084$, 95% CI 0.014-0.154; $P=.02$) and gender ($b=-3.414$, 95% CI -6.118 to -0.710; $P=.02$) on the pooled effect size (Table 2; Figure S6 in Multimedia Appendix 1).

Table 3. Meta-analysis of efficacy of conversational agent interventions for specific anxiety symptoms and quality of life.

	Specific anxiety symptoms				Quality of life			
	Value, n ^a	g (95% CI)	I ² (%; 95% CI)	P value ^b	Value, n ^a	g (95% CI)	I ² (%; 95% CI)	P value ^b
Overall effect analysis				N/A ^c				N/A
Short-term effect	18	0.47 (0.07 to 0.86) ^d	93.17 (90.62 to 95.03)		12	0.27 (0.16 to 0.39) ^e	44.16 (0.00 to 71.54)	
Adjusted for publication bias	18	0.47 (0.07 to 0.86) ^d	N/A		15	0.18 (0.07 to 0.29) ^e	N/A	
Long-term effect	14	0.11 (−0.32 to 0.55)	91.82 (88.01 to 94.42)		8	−0.04 (−0.19 to 0.11)	24.17 (0.00 to 64.28)	
Adjusted for publication bias	14	0.11 (−0.23 to 0.67)	N/A		8	−0.04 (−0.19 to 0.11)	N/A	
Sensitivity analysis				N/A				N/A
One effect size per study (largest)	6	0.88 (0.35 to 1.41) ^e	88.44 (77.37 to 94.09)		11	0.27 (0.16 to 0.39) ^e	49.21 (0.00 to 74.60)	
One effect size per study (smallest)	6	0.22 (−0.77 to 1.20)	96.64 (94.64 to 97.89)		11	0.27 (0.15 to 0.39) ^e	49.21 (0.00 to 74.60)	
Low risk of bias only (all criteria met)	5	0.54 (−0.67 to 1.75)	97.43 (95.84 to 98.40)		3	0.38 (−0.14 to 0.89)	83.54 (50.25 to 94.55)	
Subgroup analyses								
CA^f type				.09				.04 ^g
Chatbot	2	0.03 (−0.33 to 0.39)	0.00		5	0.54 (0.26 to 0.83) ^e	29.70 (0.00 to 72.79)	
Embodied CA	0	— ^h	—		2	0.21 (0.03 to 0.40) ^d	0.00	
CA in virtual reality	16	0.52 (0.08 to 0.96) ^d	93.78 (91.36 to 95.52)		2	−0.08 (−0.40 to 0.24)	0.00	
Avatar	0	—	—		3	0.24 (−0.02 to 0.50)	0.00 (0.00 to 89.60)	
Self-guided				—				.69
Yes	18	0.47 (0.07 to 0.86) ^d	93.17 (90.62 to 95.03)		9	0.31 (0.08 to 0.54) ^d	55.10 (5.08 to 78.76)	
No	0	—	—		3	0.24 (−0.02 to 0.50)	0.00 (0.00 to 89.60)	
Control condition type				<.001				.07
Waitlist or assessment only	10	0.99 (0.63 to 1.36) ^e	86.59 (77.26 to 92.09)		6	0.44 (0.16 to 0.71) ⁱ	56.67 (0.00 to 82.54)	
Information or attentional control	2	0.03 (−0.33 to 0.39)	0.00		0	—	—	
Active control	6	−0.30 (−0.73 to 0.14)	80.45 (57.78 to 90.96)		6	0.13 (−0.07 to 0.32)	0.00 (0.00 to 74.63)	
Intervention target				—				.49
Directly aimed at this outcome	18	0.47 (0.07 to 0.86) ^d	93.17 (90.62 to 95.03)		3	0.21 (0.04 to 0.39) ^d	0.00 (0.00 to 89.60)	
Not directly aimed at this outcome	0	—	—		9	0.32 (0.07 to 0.58) ^d	57.51 (10.85 to 79.75)	
Intervention length (weeks)				<.001				.36

	Specific anxiety symptoms				Quality of life			
	Value, n ^a	g (95% CI)	I ² (%; 95% CI)	P value ^b	Value, n ^a	g (95% CI)	I ² (%; 95% CI)	P value ^b
0-4	9	-0.09 (-0.46 to 0.28)	82.68 (68.48 to 90.48)		5	0.10 (-0.16 to 0.39)	11.11 (0.00 to 53.67)	
5-8	9	1.01 (0.61 to 1.42) ^e	87.81 (79.00 to 92.93)		1	0.43 (-0.44 to 1.31)	N/A	
≥9	0	—	—		6	0.37 (0.13 to 0.62) ⁱ	61.24 (5.30 to 84.14)	
Follow-up length^j (weeks)				<i><.001</i>				.54
0-8	2	1.96 (1.62 to 2.30) ^e	0.00		0	—	—	
9-16	3	-0.22 (-0.86 to 0.42)	87.02 (62.98 to 95.45)		3	0.06 (-0.33 to 0.44)	59.50 (0.00 to 88.46)	
≥17	9	-0.19 (-0.36 to -0.01) ^d	12.25 (0.00 to 53.96)		6	-0.08 (-0.28 to 0.12)	4.08 (0.00 to 32.89)	
Target sample				.52				.34
Clinical sample	8	0.45 (0.18 to 0.71) ^e	60.50 (14.22 to 81.81)		3	0.21 (-0.05 to 0.47)	0.00 (0.00 to 89.60)	
Symptomatic sample	7	0.59 (-0.31 to 1.49)	97.34 (96.05 to 98.21)		5	0.49 (0.10 to 0.88) ^d	67.48 (15.79 to 87.44)	
Nonclinical or nonsymp-tomatic sample	3	0.23 (-0.09 to 0.55)	0.00 (0.00 to 89.60)		4	0.17 (0.01 to 0.34) ^d	0.00 (0.00 to 84.69)	
Personalization and tailor				.12				.25
Yes	5	0.14 (-0.10 to 0.38)	0.00 (0.00 to 79.20)		6	0.18 (-0.04 to 0.39)	0.00 (0.00 to 74.63)	
No	13	0.59 (0.07 to 1.10) ^d	94.93 (92.84 to 96.41)		6	0.39 (0.10 to 0.67) ⁱ	65.88 (18.31 to 85.75)	
Emotional and empathic responses				<i><.001</i>				.047
Yes	2	1.96 (1.63 to 2.30) ^e	0.00		8	0.43 (0.20 to 0.66) ^e	36.53 (0.00 to 71.95)	
No	16	0.28 (-0.09 to 0.65)	91.27 (87.44 to 93.93)		4	0.14 (-0.03 to 0.30)	1.61 (0.00 to 14.13)	
Automatic reminders to engage provided				.53				.41
Yes	4	0.67 (0.03 to 1.31) ^d	90.84 (79.63 to 95.88)		3	0.42 (0.10 to 0.74) ^d	0.00 (0.00 to 89.60)	
No	14	0.41 (-0.07 to 0.89)	93.17 (90.18 to 95.25)		9	0.26 (0.04 to 0.48) ^d	55.83 (6.81 to 79.06)	
Cognitive behavioral therapy – based CA				.008				.08
Yes	6	1.09 (0.50 to 1.68) ^e	91.78 (84.88 to 95.54)		7	0.43 (0.18 to 0.69) ^e	48.28 (0.00 to 78.14)	

	Specific anxiety symptoms				Quality of life			
	Value, n ^a	<i>g</i> (95% CI)	<i>I</i> ² (%; 95% CI)	<i>P</i> value ^b	Value, n ^a	<i>g</i> (95% CI)	<i>I</i> ² (%; 95% CI)	<i>P</i> value ^b
No	12	0.15 (–0.23 to 0.53)	87.54 (80.09 to 92.20)		5	0.12 (–0.10 to 0.35)	18.03 (0.00 to 63.71)	

^aNumber of comparisons.

^b*P* value represents the significance of the *Q* test.

^cN/A: not applicable.

^d*P*<.05.

^e*P*<.001.

^fCA: conversational agent.

^gItalicized values indicate statistically significant differences.

^hMissing data.

ⁱ*P*<.01.

^jFollow-up length is a moderator of long-term effects, whereas the other variables are moderators of short-term effects.

Quality of Life or Well-being

The pooled effect size for the 12 postintervention comparisons was $g=0.27$ (95% CI 0.16 to 0.39), with moderate heterogeneity ($I^2=44.16\%$, 95% CI 0.00%-71.54%; [Figure 2](#); [Table 3](#); [Figure S11 in Multimedia Appendix 1](#)). It did not remain significant when restricting the analyses to trials with a low risk of bias ([Table 3](#)). The pooled effect size of long-term efficacy was $g=-0.04$ (95% CI –0.19 to 0.11), with low heterogeneity ($I^2=24.17\%$, 95% CI 0.00%-64.28%; [Figure 2](#); [Table 3](#); [Figure S12 in Multimedia Appendix 1](#)).

The subgroup analyses revealed 2 statistically significant moderators ([Table 3](#); [Figure S13 Multimedia Appendix 1](#)). Larger effect sizes were found in studies with chatbot ($P=.04$) and those that supported emotional and empathic responses ($P=.047$).

Meta-regression analyses revealed no statistically significant results ([Table 2](#)).

General Distress

The pooled effect size for the 12 postintervention comparisons was $g=0.33$ (95% CI 0.20-0.45), with low heterogeneity ($I^2=6.93\%$, 95% CI 0.00%-43.77%; [Figure 2](#); [Table 4](#); [Figure S14 in Multimedia Appendix 1](#)). It remained significant across all sensitivity analyses ([Table 4](#)). The pooled effect size of long-term efficacy was $g=0.39$, with low heterogeneity ($I^2=0.00\%$, 95% CI 0.00%-70.81%; [Figure 2](#); [Table 4](#); [Figure S15 in Multimedia Appendix 1](#)).

The subgroup analyses revealed 2 statistically significant moderators ([Table 4](#); [Figure S16 in Multimedia Appendix 1](#)). Larger effect sizes were found in studies that supported personalization and tailoring ($P=.002$) and that supported emotional and empathic responses ($P=.03$).

Meta-regression analyses revealed no statistically significant results ([Table 2](#)).

Table 4. Meta-analysis of efficacy of conversational agent interventions for general distress and stress.

	General distress				Stress			
	Value, n ^a	g (95% CI)	I ² (%; 95% CI)	P value ^b	Value, n ^a	g (95% CI)	I ² (%; 95% CI)	P value ^b
Overall effect analysis				N/A ^c				N/A
Short-term effect	12	0.33 (0.20 to 0.45) ^d	6.93 (0.00 to 43.77)		7	0.24 (0.08 to 0.41) ^e	39.10 (0.00 to 74.39)	
Adjusted for publication bias	15	0.27 (0.15 to 0.39) ^d	N/A		8	0.27 (0.05 to 0.50) ^f	N/A	
Long-term effect	7	0.39 (0.19 to 0.59) ^d	0.00 (0.00 to 70.81)		4	0.09 (-0.12 to 0.29)	0.00 (0.00 to 84.69)	
Adjusted for publication bias	7	0.39 (0.19 to 0.59) ^d	N/A		4	0.09 (-0.12 to 0.24)	N/A	
Sensitivity analysis				N/A				N/A
One effect size per study (largest)	8	0.31 (0.18 to 0.45) ^d	14.92 (0.00 to 57.85)		7	0.24 (0.08 to 0.41) ^e	39.10 (0.00 to 74.39)	
One effect size per study (smallest)	8	0.28 (0.15 to 0.41) ^d	0.00 (0.00 to 67.58)		7	0.24 (0.08 to 0.41) ^e	39.10 (0.00 to 74.39)	
Low risk of bias only (all criteria met)	1	0.44 (0.04 to 0.84) ^f	— ^g		2	0.33 (0.09 to 0.57) ^e	62.55 (0.00 to 91.38)	
Subgroup analyses								
CA^h type				.06				.15
Chatbot	5	0.45 (0.19 to 0.71) ^e	31.46 (0.00 to 79.20)		5	0.24 (-0.02 to 0.49)	33.65 (0.00 to 74.92)	
Embodied CA	5	0.22 (0.06 to 0.39) ^e	0.00 (0.00 to 79.20)		1	0.91 (0.21 to 1.61) ^f	—	
CA in virtual reality	0	—	—		0	—	—	
Avatar	2	0.80 (0.29 to 1.30) ^e	0.00		1	0.15 (-0.17 to 0.47)	—	
Self-guided				.06				.46
Yes	10	0.30 (0.17 to 0.42) ^e	0.00 (0.00 to 90.15)		6	0.32 (0.03 to 0.60) ^f	47.10 (0.00 to 79.04)	
No	2	0.80 (0.29 to 1.30) ^e	0.00		1	0.15 (-0.17 to 0.47)	—	
Control condition type				.77				.01 ⁱ
Passive control	5	0.40 (0.11 to 0.69) ^e	39.45 (0.00 to 77.59)		3	0.62 (0.32 to 0.91) ^d	0.00 (0.00 to 89.60)	
Information or attentional control	2	0.15 (-0.51 to 0.80)	0.00		1	0.09 (-0.50 to 0.68)	—	
Active control	5	0.39 (0.17 to 0.61) ^d	8.82 (0.00 to 48.55)		3	0.08 (-0.12 to 0.29)	0.00 (0.00 to 89.60)	
Intervention target				.11				.05
Directly aimed at this outcome	10	0.43 (0.25 to 0.60) ^d	0.52 (0.00 to 8.00)		2	0.73 (0.22 to 1.25) ^e	0.00	
Not directly aimed at this outcome	2	0.22 (0.04 to 0.40) ^f	0.00		5	0.19 (-0.02 to 0.40)	26.18 (0.00 to 70.60)	
Intervention length (weeks)				.11				.77

	General distress				Stress			
	Value, n ^a	<i>g</i> (95% CI)	<i>I</i> ² (%; 95% CI)	<i>P</i> value ^b	Value, n ^a	<i>g</i> (95% CI)	<i>I</i> ² (%; 95% CI)	<i>P</i> value ^b
0-4	9	0.44 (0.25 to 0.64) ^d	13.77 (0.00 to 55.79)		4	0.30 (-0.10 to 0.70)	51.35 (0.00 to 83.92)	
5-8	0	—	—		1	0.09 (-0.49 to 0.67)	—	
≥9	3	0.23 (0.05 to 0.41) ^f	0.00 (0.00 to 89.60)		2	0.35 (-0.05 to 0.74)	62.55 (0.00 to 91.38)	
Follow-up length^j (weeks)				—				.44
0-8	7	0.39 (0.19 to 0.59) ^d	0.00 (0.00 to 70.81)		2	-0.11 (-0.48 to 0.27)	0.00	
9-16	0	—	—		1	0.10 (-0.26 to 0.46)	—	
≥17	0	—	—		1	0.22 (-0.11 to 0.54)	—	
Target sample				.61				.51
Clinical sample	4	0.56 (0.16 to 0.95) ^e	0.00 (0.00 to 84.69)		2	0.14 (-0.14 to 0.42)	0.00	
Symptomatic sample	4	0.22 (-0.01 to 0.46)	0.00 (0.00 to 84.69)		4	0.34 (-0.04 to 0.73)	65.24 (0.00 to 88.19)	
Nonclinical or nonsymp-tomatic sample	4	0.48 (0.17 to 0.78) ^e	52.73 (0.00 to 84.69)		1	0.53 (-0.23 to 1.29)	—	
Personalization and tailor				.002				.71
Yes	4	0.82 (0.48 to 1.16) ^d	0.00 (0.00 to 84.69)		2	0.21 (-0.09 to 0.51)	0.00	
No	8	0.25 (0.12 to 0.38) ^d	0.00 (0.00 to 67.58)		5	0.29 (-0.02 to 0.61)	55.46 (0.00 to 83.54)	
Emotional and empathic responses				.03				.04
Yes	6	0.61 (0.33 to 0.89) ^d	3.68 (0.00 to 30.81)		4	0.46 (0.14 to 0.77) ^e	41.74 (0.00 to 80.38)	
No	6	0.25 (0.11 to 0.39) ^d	0.00 (0.00 to 74.63)		3	0.04 (-0.20 to 0.29)	0.00 (0.00 to 89.60)	
Automatic reminders to engage provided				.10				.46
Yes	7	0.27 (0.07 to 0.46) ^e	0.00 (0.00 to 70.81)		4	0.20 (-0.15 to 0.54)	43.32 (0.00 to 81.01)	
No	5	0.62 (0.25 to 0.98) ^e	56.69 (0.00 to 83.95)		3	0.37 (0.07 to 0.66) ^f	31.36 (0.00 to 76.73)	
Cognitive behavioral therapy – based CA				.84				.09
Yes	5	0.35 (0.10 to 0.60) ^e	15.29 (0.00 to 60.43)		3	0.44 (0.16 to 0.73) ^e	0.00 (0.00 to 89.60)	

	General distress				Stress			
	Value, n ^a	<i>g</i> (95% CI)	<i>I</i> ² (%; 95% CI)	<i>P</i> value ^b	Value, n ^a	<i>g</i> (95% CI)	<i>I</i> ² (%; 95% CI)	<i>P</i> value ^b
No	7	0.38 (0.18 to 0.58) ^d	11.07 (0.00 to 53.32)		5	0.13 (-0.10 to 0.36)	33.55 (0.00 to 74.87)	

^aNumber of comparisons.

^b*P* value represents the significance of the *Q* test.

^cN/A: not applicable.

^d*P*<.001.

^e*P*<.01.

^f*P*<.05.

^gMissing data.

^hCA: conversational agent.

ⁱItalicized values indicate statistically significant differences.

^jFollow-up length is a moderator of long-term effects, whereas the other variables are moderators of short-term effects.

Stress

The pooled effect size for the 7 postintervention comparisons was *g*=0.24 (95% CI 0.08-0.41), with moderate heterogeneity (*I*²=39.10%, 95% CI 0.00%-74.39%; Figure 2; Table 4; Figure S17 in Multimedia Appendix 1). It remained significant across all sensitivity analyses (Table 4). The pooled effect size of long-term efficacy was *g*=0.09, with low heterogeneity (*I*²=0.00%, 95% CI 0.00%-84.69%; Figure 2; Table 4; Figure S18 in Multimedia Appendix 1).

Subgroup analyses revealed 2 statistically significant differences (Table 4; Figure S19 in Multimedia Appendix 1). Larger effect sizes were found in studies with WL or assessment controls

(*P*=.01) and those that supported emotional and empathic responses (*P*=.04).

Meta-regression analyses revealed no statistically significant results (Table 2).

Other Outcomes

CAIs were significantly more effective than controls in improving mental disorder symptoms (*g*=0.36), psychosomatic disease symptoms (*g*=0.62), and negative affect (*g*=0.28), and the pooled effect sizes of long-term efficacy of mental disorder symptoms (*g*=0.31) and psychosomatic disease symptoms (*g*=0.27) were statistically significant (Figure 2; Table 5; Figures S20 and S21 in Multimedia Appendix 1).

Table 5. Meta-analysis of efficacy of conversational agent interventions for other outcomes.

Outcome measure	Value, n ^a	<i>g</i> (95% CI)	<i>I</i> ² (%; 95% CI)
Mental disorder symptoms			
Short-term effect	6	0.36 (0.17 to 0.54) ^b	0.00 (00.00 to 74.63)
Long-term effect	4	0.31 (0.07 to 0.55) ^c	0.00 (00.00 to 84.69)
Psychosomatic disease symptoms			
Short-term effect	4	0.62 (0.14 to 1.11) ^c	79.22 (44.53 to 92.21)
Long-term effect	2	0.27 (0.02 to 0.53) ^c	0.00
Positive affect			
Short-term effect	6	0.19 (-0.04 to 0.42)	0.00 (0.00 to 74.63)
Long-term effect	4	0.09 (-0.22 to 0.39)	0.00 (00.00 to 84.69)
Negative affect			
Short-term effect	6	0.28 (0.05 to 0.51) ^c	0.00 (0.00 to 74.63)
Long-term effect	4	0.16 (-0.15 to 0.46)	0.00 (00.00 to 84.69)

^aNumber of comparisons.

^b*P*<.001.

^c*P*<.05.

Discussion

Principal Findings

A total of 32 RCTs were found in our systematic review and meta-analysis to have evaluated the efficacy of CAIs in easing the symptoms of a range of mental health problems. With effect sizes ranging from $g=0.24$ to 0.62 , most of which remained robust even after performing various sensitivity analyses, they were significantly better than control conditions in improving depressive symptoms, generalized anxiety symptoms, specific anxiety symptoms, quality of life or well-being, general distress, stress, mental disorder symptoms, psychosomatic disease symptoms, and negative affect. The long-term effects of CAIs on depressive symptoms, general distress, stress, mental disorder symptoms, and psychosomatic disease symptoms were statistically significant ($g=0.16-0.39$). More high-quality evidence is needed in the future to explore the long-term efficacy of CAIs.

Different mental health problems responded well to the 4 different types of CAs. Chatbot showed the largest effect size for generalized anxiety symptoms and quality of life, ECA showed the largest effect size for depressive symptoms and stress, CA in VR showed the largest effect size for specific anxiety symptoms, and avatar showed the largest effect size for general distress. This suggested that the efficacy of the different digital methods varied.

When compared with active controls, only the CAIs for depressive symptoms and general distress showed statistically significant effect sizes. Significant effect sizes were observed in the nonsymptomatic (depressive symptoms, generalized anxiety symptoms, quality of life, and general distress), symptomatic (depressive symptoms, generalized anxiety symptoms, and quality of life), and clinical samples (depressive symptoms, specific anxiety symptoms, and general distress).

Although there was no statistically significant correlation between the length of the intervention and its efficacy, we recommend extending it to better determine how clinical outcomes and nonclinical metrics interact during the intervention of CAIs.

CAs with advanced empathetic skills have improved user affinities and experiences [77,78]. In mental health CAs, the use of user profiles or user models to support personalized and adaptive features and the assessment of personalization is still in its infancy [79]. The results of this meta-analysis demonstrate how the use of personalization and empathetic responses can significantly improve the efficacy of CAIs. In particular, empathetic responses were linked to larger effect sizes for all mental health outcomes. This suggests that future technology and mechanism research on CAs may concentrate on these 2 capacities. A breakthrough in these 2 capacities will be necessary for CAs to function as competently as human therapists.

We included automatic reminders and theoretical orientation as moderators in the subgroup analyses, as in the earlier meta-analyses of smartphone interventions. In contrast to previous research [72,80], which concluded that studies offering

engagement reminders were consistently associated with larger effect sizes, our study showed a stronger effect on depressive symptoms when the intervention did not include automatic reminders to engage. The same phenomenon was observed in the effect sizes of generalized anxiety symptoms, general distress, and stress, although it was not statistically significant. This may be because digital interventions such as internet-based CBT largely consisted of linear, structured psychotherapy modules [81,82], whereas more freedom for participants was offered in the context of CAIs [12], where interference and repetition would instead have an adverse effect on interest. We learned that the design of CAs needs to be succinct and empathic because, otherwise, the user's interest will quickly wane, which is not conducive to the establishment of a stable working alliance between CAs and participants. Coupled with feedback for CAIs of participants in some studies, such as "process violations, repetitive content, misunderstanding, impersonality, not enough interactivity, and unnatural conversation," we think that "maintaining engagement in the therapeutic process" will be the next area of focus for the development of CAIs for mental health problems rather than "attracting participation into the therapeutic process."

As found in depressive symptoms and generalized anxiety symptoms, dosage is significantly positively correlated with the efficacy of CAIs so that a moderate increase in the frequency of interaction between CAs and participants is beneficial. Future CAs work to increase participants' willingness to actively engage in conversation was supposed rather than only requiring them to accomplish a task in a passive manner.

Limitations and Future Directions

There were certain limitations in this study that should be considered. First, a thorough search was conducted, prioritizing sensitivity over specificity because of the lack of standardized language in this field. Although some of the included studies treated a specific mental health issue as a secondary outcome or auxiliary outcome, the data were still extracted and used in the corresponding meta-analysis, and we used the broadest range of participants and comparators, which may have resulted in significant heterogeneity. Second, we chose the measurement instruments that most frequently corresponded to each of the mental health outcome to extract the data. We then used the standardized mean difference to remove the effect of dimensionality; however, it should be noted that different measurement instruments for the same mental health outcome are not exactly equivalent. As a result, if sufficient research has been conducted, it makes more sense to use a single measurement tool or to perform additional subgroup analysis for the measurement tool. Finally, in meta-regression, participant characteristics can only be included as covariates at the study or trial level for analysis, which may not accurately reflect the level of individual participants, leading to *aggregate bias* [83].

Conclusions

In conclusion, our study found evidence for the efficacy of existing CAIs for mental health problems and offered the most thorough summary of their clinical traits and nonclinical metrics. When compared with various control settings and across diverse groups, CAIs significantly improved a variety of mental health

outcomes in the short term, but their long-term impacts were less than ideal. The performance and efficacy of 4 different forms of CAs (chatbot, ECA, CA in VR, and avatar) varied for various mental health problems. The efficacy of CAIs was strongly connected to 2 key facilitators: personalization and empathic response. Efficacy was not improved by receiving too many automated reminders to participate in CAIs. We also found a positive dose-response association.

In the postpandemic and digital eras, CAIs are likely to play a significant role and contribute significantly to the new health

transformation. We still require more high-quality publications, particularly evidence of direct contrasts with guided digital interventions or web-based interventions. It is intended that multidisciplinary collaboration and integration continue to advance until the divide between theoretical mechanisms and technological development is effectively eliminated. Nevertheless, it is crucial to understand that CAIs are still in their infancy and have a long way to go before they can be widely used in clinical practice and reach their full potential within existing models of mental health care.

Acknowledgments

This study was supported by the National Social Science Foundation of China (grants 14AZD111 and 21BSH017). The authors thank Xiaomi Corporation and Tianjin Quesoar Intelligent Technology Co Ltd for support on technology and theory of conversational agents; Tianjin Anding Hospital for supervision of the content for mental health; and professor Peng Zhang and associate professor Bo Wang of the College of Intelligence and Computing, Tianjin University, for their valuable suggestions for this study.

Data Availability

Deidentified data generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

YH had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. YH and LY contributed to the conception of the study. YH drafted the manuscript. CQ, QZ, XH, and LY contributed to critical revision of the manuscript for important intellectual content. YH, TL, and ZS conducted the statistical analysis. YH, CQ, XH, and LY contributed to obtaining funding. QZ and LY supervised this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The supplemental content of PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), study characteristics, risk of bias, meta-analysis, subgroup analyses, and meta-regression.

[\[PDF File \(Adobe PDF File\), 2355 KB-Multimedia Appendix 1\]](#)

References

1. World mental health report: transforming mental health for all. World Health Organization. 2022 Jun 16. URL: <https://www.who.int/publications/i/item/9789240049338> [accessed 2022-07-20]
2. Mental health and COVID-19: early evidence of the pandemic's impact: scientific brief, 2 March 2022. World Health Organization. 2022 Mar 02. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci_Brief-Mental_health-2022_1 [accessed 2022-05-30]
3. Ainiwaer A, Zhang S, Ainiwaer X, Ma F. Effects of message framing on cancer prevention and detection behaviors, intentions, and attitudes: systematic review and meta-analysis. *J Med Internet Res* 2021 Sep 16;23(9):e27634 [FREE Full text] [doi: [10.2196/27634](https://doi.org/10.2196/27634)] [Medline: [34528887](https://pubmed.ncbi.nlm.nih.gov/34528887/)]
4. Lu J, Xu X, Huang Y, Li T, Ma C, Xu G, et al. Prevalence of depressive disorders and treatment in China: a cross-sectional epidemiological study. *Lancet Psychiatry* 2021 Nov;8(11):981-990. [doi: [10.1016/S2215-0366\(21\)00251-0](https://doi.org/10.1016/S2215-0366(21)00251-0)] [Medline: [34559991](https://pubmed.ncbi.nlm.nih.gov/34559991/)]
5. Johnson S, Dalton-Locke C, Vera San Juan N, Foye U, Oram S, Papamichail A, COVID-19 Mental Health Policy Research Unit Group. Impact on mental health care and on mental health service users of the COVID-19 pandemic: a mixed methods survey of UK mental health care staff. *Soc Psychiatry Psychiatr Epidemiol* 2021 Jan;56(1):25-37 [FREE Full text] [doi: [10.1007/s00127-020-01927-4](https://doi.org/10.1007/s00127-020-01927-4)] [Medline: [32857218](https://pubmed.ncbi.nlm.nih.gov/32857218/)]
6. Adams SH, Schaub JP, Nagata JM, Park MJ, Brindis CD, Irwin Jr CE. Young adult anxiety or depressive symptoms and mental health service utilization during the COVID-19 pandemic. *J Adolesc Health* 2022 Jun;70(6):985-988 [FREE Full text] [doi: [10.1016/j.jadohealth.2022.02.023](https://doi.org/10.1016/j.jadohealth.2022.02.023)] [Medline: [35422363](https://pubmed.ncbi.nlm.nih.gov/35422363/)]
7. Kazdin AE. Annual research review: expanding mental health services through novel models of intervention delivery. *J Child Psychol Psychiatry* 2019 Apr;60(4):455-472. [doi: [10.1111/jcpp.12937](https://doi.org/10.1111/jcpp.12937)] [Medline: [29900543](https://pubmed.ncbi.nlm.nih.gov/29900543/)]

8. Christensen H. Computerised therapy for psychiatric disorders. *Lancet* 2007 Jul 14;370(9582):112-113. [doi: [10.1016/S0140-6736\(07\)61065-4](https://doi.org/10.1016/S0140-6736(07)61065-4)] [Medline: [17630022](https://pubmed.ncbi.nlm.nih.gov/17630022/)]
9. Clement S, Schauman O, Graham T, Maggioni F, Evans-Lacko S, Bezborodovs N, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. *Psychol Med* 2015 Jan;45(1):11-27. [doi: [10.1017/S0033291714000129](https://doi.org/10.1017/S0033291714000129)] [Medline: [24569086](https://pubmed.ncbi.nlm.nih.gov/24569086/)]
10. Corrigan P. How stigma interferes with mental health care. *Am Psychol* 2004 Oct;59(7):614-625. [doi: [10.1037/0003-066X.59.7.614](https://doi.org/10.1037/0003-066X.59.7.614)] [Medline: [15491256](https://pubmed.ncbi.nlm.nih.gov/15491256/)]
11. Latti EG, Stiles-Shields C, Graham AK. An overview of and recommendations for more accessible digital mental health services. *Nat Rev Psychol* 2022 Jan 26;1(2):87-100 [FREE Full text] [doi: [10.1037/0003-066X.59.7.614](https://doi.org/10.1037/0003-066X.59.7.614)]
12. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc* 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: [10.1093/jamia/ocy072](https://doi.org/10.1093/jamia/ocy072)] [Medline: [30010941](https://pubmed.ncbi.nlm.nih.gov/30010941/)]
13. Xiang YT, Yang Y, Li W, Zhang L, Zhang Q, Cheung T, et al. Timely mental health care for the 2019 novel coronavirus outbreak is urgently needed. *Lancet Psychiatry* 2020 Mar;7(3):228-229 [FREE Full text] [doi: [10.1016/S2215-0366\(20\)30046-8](https://doi.org/10.1016/S2215-0366(20)30046-8)] [Medline: [32032543](https://pubmed.ncbi.nlm.nih.gov/32032543/)]
14. Torok M, Han J, Baker S, Werner-Seidler A, Wong I, Larsen ME, et al. Suicide prevention using self-guided digital interventions: a systematic review and meta-analysis of randomised controlled trials. *Lancet Digit Health* 2020 Jan;2(1):e25-e36 [FREE Full text] [doi: [10.1016/S2589-7500\(19\)30199-2](https://doi.org/10.1016/S2589-7500(19)30199-2)] [Medline: [33328037](https://pubmed.ncbi.nlm.nih.gov/33328037/)]
15. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *Can J Psychiatry* 2019 Jul;64(7):456-464 [FREE Full text] [doi: [10.1177/0706743719828977](https://doi.org/10.1177/0706743719828977)] [Medline: [30897957](https://pubmed.ncbi.nlm.nih.gov/30897957/)]
16. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. *J Med Internet Res* 2017 May 09;19(5):e151 [FREE Full text] [doi: [10.2196/jmir.6553](https://doi.org/10.2196/jmir.6553)] [Medline: [28487267](https://pubmed.ncbi.nlm.nih.gov/28487267/)]
17. Tudor Car L, Dhinakaran DA, Kyaw BM, Kowatsch T, Joty S, Theng YL, et al. Conversational agents in health care: scoping review and conceptual analysis. *J Med Internet Res* 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: [10.2196/17158](https://doi.org/10.2196/17158)] [Medline: [32763886](https://pubmed.ncbi.nlm.nih.gov/32763886/)]
18. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. *JMIR Ment Health* 2019 Oct 18;6(10):e14166 [FREE Full text] [doi: [10.2196/14166](https://doi.org/10.2196/14166)] [Medline: [31628789](https://pubmed.ncbi.nlm.nih.gov/31628789/)]
19. Ho A, Hancock J, Miner AS. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *J Commun* 2018 Aug;68(4):712-733 [FREE Full text] [doi: [10.1093/joc/jqy026](https://doi.org/10.1093/joc/jqy026)] [Medline: [30100620](https://pubmed.ncbi.nlm.nih.gov/30100620/)]
20. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Educ Couns* 2005 Oct;59(1):21-30. [doi: [10.1016/j.pec.2004.09.008](https://doi.org/10.1016/j.pec.2004.09.008)] [Medline: [16198215](https://pubmed.ncbi.nlm.nih.gov/16198215/)]
21. Dosovitsky G, Bunge EL. Bonding with bot: user feedback on a chatbot for social isolation. *Front Digit Health* 2021 Oct 06;3:735053 [FREE Full text] [doi: [10.3389/fdgh.2021.735053](https://doi.org/10.3389/fdgh.2021.735053)] [Medline: [34713203](https://pubmed.ncbi.nlm.nih.gov/34713203/)]
22. Beatty C, Malik T, Meheli S, Sinha C. Evaluating the therapeutic alliance with a free-text CBT conversational agent (Wysa): a mixed-methods study. *Front Digit Health* 2022 Apr 11;4:847991 [FREE Full text] [doi: [10.3389/fdgh.2022.847991](https://doi.org/10.3389/fdgh.2022.847991)] [Medline: [35480848](https://pubmed.ncbi.nlm.nih.gov/35480848/)]
23. Darcy A, Daniels J, Salinger D, Wicks P, Robinson A. Evidence of human-level bonds established with a digital conversational agent: cross-sectional, retrospective observational study. *JMIR Form Res* 2021 May 11;5(5):e27868 [FREE Full text] [doi: [10.2196/27868](https://doi.org/10.2196/27868)] [Medline: [33973854](https://pubmed.ncbi.nlm.nih.gov/33973854/)]
24. He Y, Yang L, Zhu X, Wu B, Zhang S, Qian C, et al. Mental health chatbot for young adults with depressive symptoms during the COVID-19 pandemic: single-blind, three-arm randomized controlled trial. *J Med Internet Res* 2022 Nov 21;24(11):e40719 [FREE Full text] [doi: [10.2196/40719](https://doi.org/10.2196/40719)] [Medline: [36355633](https://pubmed.ncbi.nlm.nih.gov/36355633/)]
25. Xu B, Zhuang Z. Survey on psychotherapy chatbots. *Concurr Comput Pract Exp* 2022 Mar 25;34(7):e6170 [FREE Full text] [doi: [10.1002/cpe.6170](https://doi.org/10.1002/cpe.6170)]
26. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry* 2021 Oct;20(3):318-335 [FREE Full text] [doi: [10.1002/wps.20883](https://doi.org/10.1002/wps.20883)] [Medline: [34505369](https://pubmed.ncbi.nlm.nih.gov/34505369/)]
27. Craig TK, Rus-Calafell M, Ward T, Leff JP, Huckvale M, Howarth E, et al. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. *Lancet Psychiatry* 2018 Jan;5(1):31-40 [FREE Full text] [doi: [10.1016/S2215-0366\(17\)30427-3](https://doi.org/10.1016/S2215-0366(17)30427-3)] [Medline: [29175276](https://pubmed.ncbi.nlm.nih.gov/29175276/)]
28. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Ment Health* 2017 Jun 06;4(2):e19 [FREE Full text] [doi: [10.2196/mental.7785](https://doi.org/10.2196/mental.7785)] [Medline: [28588005](https://pubmed.ncbi.nlm.nih.gov/28588005/)]
29. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. *JMIR Ment Health* 2018 Dec 13;5(4):e64 [FREE Full text] [doi: [10.2196/mental.9782](https://doi.org/10.2196/mental.9782)] [Medline: [30545815](https://pubmed.ncbi.nlm.nih.gov/30545815/)]

30. Greer S, Ramo D, Chang YJ, Fu M, Moskowitz J, Haritatos J. Use of the Chatbot "Vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial. *JMIR Mhealth Uhealth* 2019 Oct 31;7(10):e15018 [[FREE Full text](#)] [doi: [10.2196/15018](https://doi.org/10.2196/15018)] [Medline: [31674920](https://pubmed.ncbi.nlm.nih.gov/31674920/)]
31. Jang S, Kim JJ, Kim SJ, Hong J, Kim S, Kim E. Mobile app-based chatbot to deliver cognitive behavioral therapy and psychoeducation for adults with attention deficit: a development and feasibility/usability study. *Int J Med Inform* 2021 Jun;150:104440. [doi: [10.1016/j.ijmedinf.2021.104440](https://doi.org/10.1016/j.ijmedinf.2021.104440)] [Medline: [33799055](https://pubmed.ncbi.nlm.nih.gov/33799055/)]
32. Oh J, Jang S, Kim H, Kim JJ. Efficacy of mobile app-based interactive cognitive behavioral therapy using a chatbot for panic disorder. *Int J Med Inform* 2020 Aug;140:104171. [doi: [10.1016/j.ijmedinf.2020.104171](https://doi.org/10.1016/j.ijmedinf.2020.104171)] [Medline: [32446158](https://pubmed.ncbi.nlm.nih.gov/32446158/)]
33. Bennion MR, Hardy GE, Moore RK, Kellett S, Millings A. Usability, acceptability, and effectiveness of web-based conversational agents to facilitate problem solving in older adults: controlled study. *J Med Internet Res* 2020 May 27;22(5):e16794 [[FREE Full text](#)] [doi: [10.2196/16794](https://doi.org/10.2196/16794)] [Medline: [32384055](https://pubmed.ncbi.nlm.nih.gov/32384055/)]
34. Bird T, Mansell W, Wright J, Gaffney H, Tai S. Manage your life online: a web-based randomized controlled trial evaluating the effectiveness of a problem-solving intervention in a student sample. *Behav Cogn Psychother* 2018 Sep;46(5):570-582 [[FREE Full text](#)] [doi: [10.1017/S1352465817000820](https://doi.org/10.1017/S1352465817000820)] [Medline: [29366432](https://pubmed.ncbi.nlm.nih.gov/29366432/)]
35. Gaffney H, Mansell W, Edwards R, Wright J. Manage your life online (MYLO): a pilot trial of a conversational computer-based intervention for problem solving in a student sample. *Behav Cogn Psychother* 2014 Nov;42(6):731-746. [doi: [10.1017/S135246581300060X](https://doi.org/10.1017/S135246581300060X)] [Medline: [23899405](https://pubmed.ncbi.nlm.nih.gov/23899405/)]
36. Ly KH, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. *Internet Interv* 2017 Oct 10;10:39-46 [[FREE Full text](#)] [doi: [10.1016/j.invent.2017.10.002](https://doi.org/10.1016/j.invent.2017.10.002)] [Medline: [30135751](https://pubmed.ncbi.nlm.nih.gov/30135751/)]
37. Prochaska JJ, Vogel EA, Chieng A, Baiocchi M, Maglalang DD, Pajarito S, et al. A randomized controlled trial of a therapeutic relational agent for reducing substance misuse during the COVID-19 pandemic. *Drug Alcohol Depend* 2021 Oct 01;227:108986 [[FREE Full text](#)] [doi: [10.1016/j.drugalcdep.2021.108986](https://doi.org/10.1016/j.drugalcdep.2021.108986)] [Medline: [34507061](https://pubmed.ncbi.nlm.nih.gov/34507061/)]
38. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. *JMIR Mhealth Uhealth* 2020 Apr 03;8(4):e15806 [[FREE Full text](#)] [doi: [10.2196/15806](https://doi.org/10.2196/15806)] [Medline: [32242820](https://pubmed.ncbi.nlm.nih.gov/32242820/)]
39. Lavelle J, Dunne N, Mulcahy HE, McHugh L. Chatbot-delivered cognitive defusion versus cognitive restructuring for negative self-referential thoughts: a pilot study. *Psychol Rec* 2021 Aug 24;72(2):247-261 [[FREE Full text](#)] [doi: [10.1007/s40732-021-00478-7](https://doi.org/10.1007/s40732-021-00478-7)]
40. Maeda E, Miyata A, Boivin J, Nomura K, Kumazawa Y, Shirasawa H, et al. Promoting fertility awareness and preconception health using a chatbot: a randomized controlled trial. *Reprod Biomed Online* 2020 Dec;41(6):1133-1143. [doi: [10.1016/j.rbmo.2020.09.006](https://doi.org/10.1016/j.rbmo.2020.09.006)] [Medline: [33039321](https://pubmed.ncbi.nlm.nih.gov/33039321/)]
41. Hunt M, Miguez S, Dukas B, Onwude O, White S. Efficacy of zemydy, a mobile digital therapeutic for the self-management of irritable bowel syndrome: crossover randomized controlled trial. *JMIR Mhealth Uhealth* 2021 May 20;9(5):e26152 [[FREE Full text](#)] [doi: [10.2196/26152](https://doi.org/10.2196/26152)] [Medline: [33872182](https://pubmed.ncbi.nlm.nih.gov/33872182/)]
42. Klos MC, Escoredo M, Joerin A, Lemos VN, Rauws M, Bunge EL. Artificial intelligence-based chatbot for anxiety and depression in university students: pilot randomized controlled trial. *JMIR Form Res* 2021 Aug 12;5(8):e20678 [[FREE Full text](#)] [doi: [10.2196/20678](https://doi.org/10.2196/20678)] [Medline: [34092548](https://pubmed.ncbi.nlm.nih.gov/34092548/)]
43. Liu H, Peng H, Song X, Xu C, Zhang M. Using AI chatbots to provide self-help depression interventions for university students: a randomized trial of effectiveness. *Internet Interv* 2022 Jan 06;27:100495 [[FREE Full text](#)] [doi: [10.1016/j.invent.2022.100495](https://doi.org/10.1016/j.invent.2022.100495)] [Medline: [35059305](https://pubmed.ncbi.nlm.nih.gov/35059305/)]
44. Ali R, Hoque E, Duberstein P, Schubert L, Razavi SZ, Kane B, et al. Aging and engaging: a pilot randomized controlled trial of an online conversational skills coach for older adults. *Am J Geriatr Psychiatry* 2021 Aug;29(8):804-815 [[FREE Full text](#)] [doi: [10.1016/j.jagp.2020.11.004](https://doi.org/10.1016/j.jagp.2020.11.004)] [Medline: [33308893](https://pubmed.ncbi.nlm.nih.gov/33308893/)]
45. Burton C, Szentagotai Tatar A, McKinstry B, Matheson C, Matu S, Moldovan R, Help4Mood Consortium. Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression. *J Telemed Telecare* 2016 Sep;22(6):348-355. [doi: [10.1177/1357633X15609793](https://doi.org/10.1177/1357633X15609793)] [Medline: [26453910](https://pubmed.ncbi.nlm.nih.gov/26453910/)]
46. Suganuma S, Sakamoto D, Shimoyama H. An embodied conversational agent for unguided internet-based cognitive behavior therapy in preventative mental health: feasibility and acceptability pilot trial. *JMIR Ment Health* 2018 Jul 31;5(3):e10454 [[FREE Full text](#)] [doi: [10.2196/10454](https://doi.org/10.2196/10454)] [Medline: [30064969](https://pubmed.ncbi.nlm.nih.gov/30064969/)]
47. Loveys K, Sagar M, Pickering I, Broadbent E. A digital human for delivering a remote loneliness and stress intervention to at-risk younger and older adults during the COVID-19 pandemic: randomized pilot trial. *JMIR Ment Health* 2021 Nov 08;8(11):e31586 [[FREE Full text](#)] [doi: [10.2196/31586](https://doi.org/10.2196/31586)] [Medline: [34596572](https://pubmed.ncbi.nlm.nih.gov/34596572/)]
48. Sandoval LR, Buckey JC, Ainslie R, Tombari M, Stone W, Hegel MT. Randomized controlled trial of a computerized interactive media-based problem solving treatment for depression. *Behav Ther* 2017 May;48(3):413-425 [[FREE Full text](#)] [doi: [10.1016/j.beth.2016.04.001](https://doi.org/10.1016/j.beth.2016.04.001)] [Medline: [28390503](https://pubmed.ncbi.nlm.nih.gov/28390503/)]
49. Cartreine JA, Locke SE, Buckey JC, Sandoval L, Hegel MT. Electronic problem-solving treatment: description and pilot study of an interactive media treatment for depression. *JMIR Res Protoc* 2012 Sep 25;1(2):e11 [[FREE Full text](#)] [doi: [10.2196/resprot.1925](https://doi.org/10.2196/resprot.1925)] [Medline: [23611902](https://pubmed.ncbi.nlm.nih.gov/23611902/)]

50. Freeman D, Haselton P, Freeman J, Spanlang B, Kishore S, Albery E, et al. Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. *Lancet Psychiatry* 2018 Aug;5(8):625-632 [FREE Full text] [doi: [10.1016/S2215-0366\(18\)30226-8](https://doi.org/10.1016/S2215-0366(18)30226-8)] [Medline: [30007519](https://pubmed.ncbi.nlm.nih.gov/30007519/)]
51. Miloff A, Lindner P, Dafgård P, Deak S, Garke M, Hamilton W, et al. Automated virtual reality exposure therapy for spider phobia vs. in-vivo one-session treatment: a randomized non-inferiority trial. *Behav Res Ther* 2019 Jul;118:130-140. [doi: [10.1016/j.brat.2019.04.004](https://doi.org/10.1016/j.brat.2019.04.004)] [Medline: [31075675](https://pubmed.ncbi.nlm.nih.gov/31075675/)]
52. Bentz D, Wang N, Ibach MK, Schicktanz NS, Zimmer A, Papassotiropoulos A, et al. Effectiveness of a stand-alone, smartphone-based virtual reality exposure app to reduce fear of heights in real-life: a randomized trial. *NPJ Digit Med* 2021 Feb 08;4(1):16 [FREE Full text] [doi: [10.1038/s41746-021-00387-7](https://doi.org/10.1038/s41746-021-00387-7)] [Medline: [33558625](https://pubmed.ncbi.nlm.nih.gov/33558625/)]
53. Donker T, Cornelisz I, van Klaveren C, van Straten A, Carlbring P, Cuijpers P, et al. Effectiveness of self-guided app-based virtual reality cognitive behavior therapy for acrophobia: a randomized clinical trial. *JAMA Psychiatry* 2019 Jul 01;76(7):682-690 [FREE Full text] [doi: [10.1001/jamapsychiatry.2019.0219](https://doi.org/10.1001/jamapsychiatry.2019.0219)] [Medline: [30892564](https://pubmed.ncbi.nlm.nih.gov/30892564/)]
54. Lindner P, Miloff A, Fagnäs S, Andersen J, Sigeman M, Andersson G, et al. Therapist-led and self-led one-session virtual reality exposure therapy for public speaking anxiety with consumer hardware and software: a randomized controlled trial. *J Anxiety Disord* 2019 Jan;61:45-54. [doi: [10.1016/j.janxdis.2018.07.003](https://doi.org/10.1016/j.janxdis.2018.07.003)] [Medline: [30054173](https://pubmed.ncbi.nlm.nih.gov/30054173/)]
55. Dellazizzo L, Potvin S, Phraxayavong K, Dumais A. One-year randomized trial comparing virtual reality-assisted therapy to cognitive-behavioral therapy for patients with treatment-resistant schizophrenia. *NPJ Schizophr* 2021 Feb 12;7(1):9 [FREE Full text] [doi: [10.1038/s41537-021-00139-2](https://doi.org/10.1038/s41537-021-00139-2)] [Medline: [33580033](https://pubmed.ncbi.nlm.nih.gov/33580033/)]
56. Kocur M, Dechant M, Wolff C, Nothdurfter C, Wetter TC, Rupprecht R, et al. Computer-assisted avatar-based treatment for dysfunctional beliefs in depressive inpatients: a pilot study. *Front Psychiatry* 2021 Jul 15;12:608997 [FREE Full text] [doi: [10.3389/fpsy.2021.608997](https://doi.org/10.3389/fpsy.2021.608997)] [Medline: [34335319](https://pubmed.ncbi.nlm.nih.gov/34335319/)]
57. Pinto MD, Hickman Jr RL, Clochesy J, Buchner M. Avatar-based depression self-management technology: promising approach to improve depressive symptoms among young adults. *Appl Nurs Res* 2013 Feb;26(1):45-48 [FREE Full text] [doi: [10.1016/j.apnr.2012.08.003](https://doi.org/10.1016/j.apnr.2012.08.003)] [Medline: [23265918](https://pubmed.ncbi.nlm.nih.gov/23265918/)]
58. du Sert OP, Potvin S, Lipp O, Dellazizzo L, Laurelli M, Breton R, et al. Virtual reality therapy for refractory auditory verbal hallucinations in schizophrenia: a pilot clinical trial. *Schizophr Res* 2018 Jul;197:176-181. [doi: [10.1016/j.schres.2018.02.031](https://doi.org/10.1016/j.schres.2018.02.031)] [Medline: [29486956](https://pubmed.ncbi.nlm.nih.gov/29486956/)]
59. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009 Jul 21;339:b2535 [FREE Full text] [doi: [10.1136/bmj.b2535](https://doi.org/10.1136/bmj.b2535)] [Medline: [19622551](https://pubmed.ncbi.nlm.nih.gov/19622551/)]
60. McTear MF, Callejas Z, Griol D. *The Conversational Interface: Talking to Smart Devices*. Cham, Switzerland: Springer; May 19, 2016.
61. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun ACM* 1966 Jan;9(1):36-45 [FREE Full text] [doi: [10.1145/365153.365168](https://doi.org/10.1145/365153.365168)]
62. Scholten MR, Kelders SM, Van Gemert-Pijnen JE. Self-guided web-based interventions: scoping review on user needs and the potential of embodied conversational agents to address them. *J Med Internet Res* 2017 Nov 16;19(11):e383 [FREE Full text] [doi: [10.2196/jmir.7351](https://doi.org/10.2196/jmir.7351)] [Medline: [29146567](https://pubmed.ncbi.nlm.nih.gov/29146567/)]
63. Ma T, Sharifi H, Chattopadhyay D. Virtual humans in health-related interventions: a meta-analysis. In: *Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019 May Presented at: CHI EA '19; May 4-9, 2019; Glasgow, Scotland p. 1-6 URL: <https://dl.acm.org/doi/proceedings/10.1145/3290607> [doi: [10.1145/3290607.3312853](https://doi.org/10.1145/3290607.3312853)]
64. Loveys K, Sagar M, Broadbent E. The effect of multimodal emotional expression on responses to a digital human during a self-disclosure conversation: a computational analysis of user language. *J Med Syst* 2020 Jul 22;44(9):143. [doi: [10.1007/s10916-020-01624-4](https://doi.org/10.1007/s10916-020-01624-4)] [Medline: [32700060](https://pubmed.ncbi.nlm.nih.gov/32700060/)]
65. Gris I, Rivera DA, Rayon A, Camacho A, Novick DG. Young merlin: an embodied conversational agent in virtual reality. In: *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. 2016 Presented at: ICMI '16; November 12-16, 2016; Tokyo, Japan p. 425-426 URL: <https://dl.acm.org/doi/10.1145/2993148.2998534> [doi: [10.1145/2993148.2998534](https://doi.org/10.1145/2993148.2998534)]
66. McVay MA, Bennett GG, Steinberg D, Voils CI. Dose-response research in digital health interventions: concepts, considerations, and challenges. *Health Psychol* 2019 Dec;38(12):1168-1174 [FREE Full text] [doi: [10.1037/hea0000805](https://doi.org/10.1037/hea0000805)] [Medline: [31580127](https://pubmed.ncbi.nlm.nih.gov/31580127/)]
67. Gal R, May AM, van Overmeeren EJ, Simons M, Monnikhof EM. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis. *Sports Med Open* 2018 Sep 03;4(1):42 [FREE Full text] [doi: [10.1186/s40798-018-0157-9](https://doi.org/10.1186/s40798-018-0157-9)] [Medline: [30178072](https://pubmed.ncbi.nlm.nih.gov/30178072/)]
68. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. *World Psychiatry* 2019 Oct;18(3):325-336 [FREE Full text] [doi: [10.1002/wps.20673](https://doi.org/10.1002/wps.20673)] [Medline: [31496095](https://pubmed.ncbi.nlm.nih.gov/31496095/)]
69. Higgins JP, Thompson SG. Controlling the risk of spurious findings from meta-regression. *Stat Med* 2004 Jun 15;23(11):1663-1682. [doi: [10.1002/sim.1752](https://doi.org/10.1002/sim.1752)] [Medline: [15160401](https://pubmed.ncbi.nlm.nih.gov/15160401/)]

70. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011 Oct 18;343:d5928 [[FREE Full text](#)] [doi: [10.1136/bmj.d5928](https://doi.org/10.1136/bmj.d5928)] [Medline: [22008217](https://pubmed.ncbi.nlm.nih.gov/22008217/)]
71. Hedges LV, Olkin I. *Statistical Methods for Meta-analysis*. New York, NY, USA: Academic Press; Jul 10, 1985.
72. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003 Sep 06;327(7414):557-560 [[FREE Full text](#)] [doi: [10.1136/bmj.327.7414.557](https://doi.org/10.1136/bmj.327.7414.557)] [Medline: [12958120](https://pubmed.ncbi.nlm.nih.gov/12958120/)]
73. Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. *Psychol Methods* 1998 Dec;3(4):486-504 [[FREE Full text](#)] [doi: [10.1037/1082-989x.3.4.486](https://doi.org/10.1037/1082-989x.3.4.486)]
74. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994 Dec;50(4):1088-1101. [Medline: [7786990](https://pubmed.ncbi.nlm.nih.gov/7786990/)]
75. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997 Sep 13;315(7109):629-634 [[FREE Full text](#)] [doi: [10.1136/bmj.315.7109.629](https://doi.org/10.1136/bmj.315.7109.629)] [Medline: [9310563](https://pubmed.ncbi.nlm.nih.gov/9310563/)]
76. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000 Jun;56(2):455-463. [doi: [10.1111/j.0006-341x.2000.00455.x](https://doi.org/10.1111/j.0006-341x.2000.00455.x)] [Medline: [10877304](https://pubmed.ncbi.nlm.nih.gov/10877304/)]
77. de Gennaro M, Krumhuber EG, Lucas G. Effectiveness of an empathic chatbot in combating adverse effects of social exclusion on mood. *Front Psychol* 2020 Jan 23;10:3061 [[FREE Full text](#)] [doi: [10.3389/fpsyg.2019.03061](https://doi.org/10.3389/fpsyg.2019.03061)] [Medline: [32038415](https://pubmed.ncbi.nlm.nih.gov/32038415/)]
78. Morris RR, Kouddous K, Kshirsagar R, Schueller SM. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. *J Med Internet Res* 2018 Jun 26;20(6):e10148 [[FREE Full text](#)] [doi: [10.2196/10148](https://doi.org/10.2196/10148)] [Medline: [29945856](https://pubmed.ncbi.nlm.nih.gov/29945856/)]
79. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: systematic review. *J Med Internet Res* 2019 Nov 07;21(11):e15360 [[FREE Full text](#)] [doi: [10.2196/15360](https://doi.org/10.2196/15360)] [Medline: [31697237](https://pubmed.ncbi.nlm.nih.gov/31697237/)]
80. White V, Linardon J, Stone JE, Holmes-Truscott E, Olive L, Mikocka-Walus A, et al. Online psychological interventions to reduce symptoms of depression, anxiety, and general distress in those with chronic health conditions: a systematic review and meta-analysis of randomized controlled trials. *Psychol Med* 2022 Feb;52(3):548-573. [doi: [10.1017/S0033291720002251](https://doi.org/10.1017/S0033291720002251)] [Medline: [32674747](https://pubmed.ncbi.nlm.nih.gov/32674747/)]
81. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. *J Med Internet Res* 2019 Jul 22;21(7):e12869 [[FREE Full text](#)] [doi: [10.2196/12869](https://doi.org/10.2196/12869)] [Medline: [31333198](https://pubmed.ncbi.nlm.nih.gov/31333198/)]
82. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. *Behav Res Ther* 2017 Jan;88:19-25 [[FREE Full text](#)] [doi: [10.1016/j.brat.2016.08.012](https://doi.org/10.1016/j.brat.2016.08.012)] [Medline: [28110672](https://pubmed.ncbi.nlm.nih.gov/28110672/)]
83. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002 Jun 15;21(11):1559-1573. [doi: [10.1002/sim.1187](https://doi.org/10.1002/sim.1187)] [Medline: [12111920](https://pubmed.ncbi.nlm.nih.gov/12111920/)]

Abbreviations

- CA:** conversational agent
- CAI:** conversational agent intervention
- CBT:** cognitive behavioral therapy
- ECA:** embodied conversational agent
- PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
- RCT:** randomized controlled trial
- VR:** virtual reality
- WL:** waitlist

Edited by A Mavragani; submitted 27.10.22; peer-reviewed by B Jack, W Mansell, E Bunge; comments to author 03.02.23; revised version received 17.02.23; accepted 10.03.23; published 28.04.23

Please cite as:

He Y, Yang L, Qian C, Li T, Su Z, Zhang Q, Hou X

Conversational Agent Interventions for Mental Health Problems: Systematic Review and Meta-analysis of Randomized Controlled Trials

J Med Internet Res 2023;25:e43862

URL: <https://www.jmir.org/2023/1/e43862>

doi: [10.2196/43862](https://doi.org/10.2196/43862)

PMID: [37115595](https://pubmed.ncbi.nlm.nih.gov/37115595/)

©Yuhao He, Li Yang, Chunlian Qian, Tong Li, Zhengyuan Su, Qiang Zhang, Xiangqing Hou. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 28.04.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.