International Platform of Registered Systematic Review and Meta-analysis Protocols

INPLASY

AI-Assisted Digital Mental Health Interventions for Anxiety and Depression in Cancer Patients: A Systematic Review and Meta-Analysis

INPLASY202610043
doi: 10.37766/inplasy2026.1.0043
Received: 13 January 2026
Published: 13 January 2026

Peng, M; Zhu, KR; Yu, JZ; Liu, XY; Liu, JR; Wang, YY; Zhang, YP; Niu, YH.

Corresponding author:
Yaohong Niu

zhangyp_789@163.com

Author Affiliation:
Xi'an Jiaotong University.

ADMINISTRATIVE INFORMATION

Support - No.

Review Stage at time of this submission - Completed but not published.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY202610043

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 13 January 2026 and was last updated on 13 January 2026.

INTRODUCTION

Review question / Objective To evaluate the effectiveness of AI-assisted digital mental health interventions for anxiety and depression in adult cancer patients.

Condition being studied Anxiety and depression are highly prevalent among cancer patients, significantly impacting treatment adherence and quality of life. Artificial intelligence (AI)-assisted digital mental health interventions have emerged as potentially scalable solutions to address the growing gap in psychological care. However, their effectiveness in cancer populations remains unclear.

METHODS

Participant or population Adult patients (aged ≥18 years) with any type of cancer diagnosis at any disease stage, including those receiving active treatment, in follow-up care, or cancer survivors.

Intervention (2) Digital interventions incorporating AI technology components, including but not limited to machine learning algorithms, natural language processing, large language models (e.g., GPT-based systems), or conversational AI agents. For the purposes of this review, we defined AI-assisted interventions as those utilizing computational systems that demonstrate adaptive behavior through learning from data, processing natural language input, or generating contextually appropriate responses beyond simple keyword matching. This definition encompasses three main categories: (a) generative AI systems using large language models (e.g., ChatGPT) that generate novel responses; (b) machine learning platforms that adapt content delivery based on user data patterns; and (c) conversational agents employing natural language processing for intent recognition, even when using predominantly scripted response libraries. We acknowledge that rule-based chatbots occupy a borderline category; we included them when they employed NLP

techniques for user input interpretation, while recognizing this represents a more limited form of AI capability. The AI component must play an active role in content delivery, personalization, or user interaction beyond simple automation. Interventions could be delivered via mobile applications, web platforms, or messaging systems.
Comparisons: (3) Control groups receiving standard care, usual care, waitlist control, or non-AI digital interventions.

Comparator Randomized controlled trials (RCTs) or quasi-experimental studies with concurrent comparison groups.

Study designs to be included Randomized controlled trials (RCTs) or quasi-experimental studies with concurrent comparison groups.retrospective comparative studies.

Eligibility criteria Eligibility Criteria
Studies were considered eligible for inclusion if they met the following criteria using the PICOS framework:
Participants: (1) Adult patients (aged ≥18 years) with any type of cancer diagnosis at any disease stage, including those receiving active treatment, in follow-up care, or cancer survivors.
Interventions: (2) Digital interventions incorporating AI technology components, including but not limited to machine learning algorithms, natural language processing, large language models (e.g., GPT-based systems), or conversational AI agents. For the purposes of this review, we defined AI-assisted interventions as those utilizing computational systems that demonstrate adaptive behavior through learning from data, processing natural language input, or generating contextually appropriate responses beyond simple keyword matching. This definition encompasses three main categories: (a) generative AI systems using large language models (e.g., ChatGPT) that generate novel responses; (b) machine learning platforms that adapt content delivery based on user data patterns; and (c) conversational agents employing natural language processing for intent recognition, even when using predominantly scripted response libraries. We acknowledge that rule-based chatbots occupy a borderline category; we included them when they employed NLP techniques for user input interpretation, while recognizing this represents a more limited form of AI capability. The AI component must play an active role in content delivery, personalization, or user interaction beyond simple automation. Interventions could be delivered via mobile applications, web platforms, or messaging systems.
Comparisons: (3) Control groups receiving standard care, usual care, waitlist control, or non-AI digital interventions.
Outcomes: (4) Studies must report quantitative data on anxiety and/or depression using validated, standardized measurement instruments such as the Hospital Anxiety and Depression Scale (HADS), Patient Health Questionnaire (PHQ-9), Generalized Anxiety Disorder scale (GAD-7), PROMIS emotional distress measures, or equivalent validated tools. Studies reporting only qualitative outcomes or non-standardized measures were excluded.
Study Design: (5) Randomized controlled trials (RCTs) or quasi-experimental studies with concurrent comparison groups. For clarity, we defined quasi-experimental studies as non-randomized controlled studies with parallel comparison groups assessed at the same time points. Pre-post studies without concurrent controls were excluded from the main analysis due to their higher susceptibility to confounding and regression to the mean, though we acknowledge their potential value for hypothesis generation.
Publication Type: (6) Full-text peer-reviewed journal articles published in English or Chinese. Conference abstracts, dissertations, and unpublished studies were excluded unless full manuscripts were available.

Information sources Electronic Database Searches
A comprehensive literature search was conducted across multiple electronic databases from January 1, 2015, to December 31, 2025, thereby synthesizing evidence from the most recent decade of research in this emerging field. The databases searched included: (1) PubMed/MEDLINE, (2) Web of Science Core Collection, (3) Google Scholar, and (4) China National Knowledge Infrastructure (CNKI) for Chinese-language publications. The search strategy combined three key concept groups using Boolean operators: (1) artificial intelligence and digital health terms ("artificial intelligence" OR "AI chatbot" OR "conversational agent" OR "virtual therapist" OR "digital mental health" OR "machine learning" OR "large language model" OR "natural language processing"), (2) cancer terms ("cancer" OR "oncology" OR "tumor" OR "neoplasm" OR "carcinoma" OR "malignancy"), and (3) mental health outcomes ("anxiety" OR "depression" OR "psychological distress" OR "mental health"). For Chinese databases, equivalent Chinese search terms were employed. The search strategy was adapted for each database according to its

specific indexing structure and controlled vocabulary. The complete search strategy for each database is provided in Supplementary Material S1.

Supplementary Search Methods

To ensure comprehensive coverage, we conducted supplementary searches including: (1) manual screening of reference lists from included studies and relevant systematic reviews, (2) forward citation tracking of key articles using Google Scholar, and (3) searching gray literature sources including conference proceedings and preprint servers. No language restrictions were imposed during the initial search phase.

Main outcome(s) (4) Studies must report quantitative data on anxiety and/or depression using validated, standardized measurement instruments such as the Hospital Anxiety and Depression Scale (HADS), Patient Health Questionnaire (PHQ-9), Generalized Anxiety Disorder scale (GAD-7), PROMIS emotional distress measures, or equivalent validated tools. Studies reporting only qualitative outcomes or non-standardized measures were excluded.

Quality assessment / Risk of bias analysis The methodological quality of included RCTs was assessed independently by two reviewers using the Cochrane Risk of Bias tool version 2 (RoB 2) [22]. This tool evaluates bias across five domains: (1) bias arising from the randomization process, (2) bias due to deviations from intended interventions, (3) bias due to missing outcome data, (4) bias in measurement of the outcome, and (5) bias in selection of the reported result. Each domain was rated as "low risk," "some concerns," or "high risk" of bias. An overall risk of bias judgment was generated for each study based on the domain-level assessments according to RoB 2 guidelines: studies were judged as "low risk" if all domains were rated as low risk, "some concerns" if at least one domain had some concerns but no domains were high risk, and "high risk" if any domain was rated as high risk. For the single quasi-experimental study, the Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I) tool was employed [23,24]. Disagreements in risk of bias assessments were resolved through discussion or consultation with a third reviewer. A summary plot and traffic light plot were generated to visualize the risk of bias across studies and domains.

Strategy of data synthesis For continuous anxiety and depression outcomes, we calculated Hedges' g as the standardized mean difference based on post-intervention data, with negative values indicating improvement favoring AI interventions and effect sizes interpreted as small (g=0.2), moderate (g=0.5), or large (g=0.8). Random-effects models using restricted maximum likelihood were employed for all meta-analyses due to expected heterogeneity across studies in participant characteristics, cancer types, AI technologies, and intervention durations, with pooled effects calculated separately for anxiety and depression and visualized through forest plots. Heterogeneity was assessed using Cochran's Q test ($p<0.10$ to increase sensitivity for detecting heterogeneity given the small number of studies), I^2 statistic (75% considerable), τ^2, and 95% prediction intervals.

Subgroup analysis Prespecified subgroup analyses explored heterogeneity by AI technology type, cancer type, and intervention duration using mixed-effects models with significance at $p<0.10$ (to avoid missing potentially important moderators in this small meta-analysis). Given the small number of studies per subgroup, we emphasize that these analyses are exploratory and intended for hypothesis generation rather than definitive conclusions.

Publication bias was evaluated through funnel plots, Egger's test ($p<0.10$ to maximize detection of small-study effects despite limited power), and small-study effects assessment, though limited study numbers ($k<10$) reduced statistical power. Cumulative meta-analysis was performed by sequentially adding studies in chronological order by publication year to examine the evolution of the pooled effect estimate over time and assess the stability of findings as evidence accumulated. All analyses were conducted using R version 4.3.0 with metafor, meta, and dmetar packages, with statistical significance set at $\alpha=0.05$ two-tailed.

Sensitivity analysis Sensitivity analyses were conducted using the leave-one-out method, whereby each study was systematically removed from the meta-analysis one at a time to assess the influence of individual studies on the pooled effect estimate and heterogeneity. This approach identifies studies that substantially alter the overall findings and helps evaluate the robustness of the meta-analytic results. For each iteration, we recalculated the pooled effect size, 95% confidence interval, and I^2 statistic. Studies were considered influential if their removal resulted in (1) a change in statistical significance of the pooled effect, (2) a shift in effect size magnitude exceeding 20%, or (3) a substantial reduction in heterogeneity ($\Delta I^2 > 25\%$).

Country(ies) involved China.

Keywords Artificial intelligence; Cancer; Anxiety; Depression; Meta-analysis.

Contributions of each author

Author 1 - Min Peng.
Author 2 - Ke-Run Zhu.
Author 3 - Jinzhuo Yu.
Author 4 - Xinyu Liu.
Author 5 - Jiarui Liu.
Author 6 - Yuyan Wang.
Author 7 - Yin-Ping Zhang.
Author 8 - Yaohong Niu.