INPLASY

Clinical and economic impact of AI mental health chatbots for older adults: systematic review and meta-analysis protocol

INPLASY202610020

doi: 10.37766/inplasy2026.1.0020

Received: 6 January 2026

Published: 6 January 2026

Balasubramanian, M; Sultana, S; Rasel, S; Shafei, A; Shivashankar, M; Reynolds, K; Curley, M; Schultz, T.

Corresponding author:
Madhan Balasubramanian

madhan.balasubramanian@flinders.edu.au

Author Affiliation:
Flinders University, College of Business Government and Law, Health Care Management | Centre for Social Impact | Factory of Future.

ADMINISTRATIVE INFORMATION

Support - Factory of Future - Strategic Industry Research Program Grant (2025).

Review Stage at time of this submission - Piloting of the study selection process.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY202610020

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 6 January 2026 and was last updated on 25 January 2026.

INTRODUCTION

Review question / Objective This review will address the overarching question: "What is the clinical effectiveness and economic impact of AI chatbot interventions that provide mental health support to older adults (≥60 years) across any setting?" The primary objective is to quantitatively estimate the effects of AI chatbots on mental health outcomes—including symptoms of depression, anxiety, psychological distress, and positive wellbeing—in older adults using data from randomised controlled trials, quasi-experimental studies, pre–post designs, and single-arm trials. A secondary objective is to synthesise available economic evidence on these interventions, including costs, resource use, benefits, and any reported or derivable cost-effectiveness metrics, and to conduct structured economic impact analyses using narrative and dominance-matrix approaches informed by previous reviews of mental health prevention and promotion. Together, this review will inform clinicians, policymakers, and technology developers about the potential value and limitations of AI chatbots as scalable solutions for older adults.

Background:
The global population of older adults is rapidly growing, from 1.1 billion in 2023 to a projected 2.1 billion by 2050, surpassing younger age groups by the late 2060s (WHO, 2024). Mental health conditions, affect ~14% of adults aged 70+ globally, significantly contributing to disability and reduced quality of life (WHO, 2024). Suicide rates in this age group are disproportionately high, accounting for ~16.6% of global suicide deaths annually (WHO, 2024). Older adults face multiple mental health risk factors, including social isolation (affecting ~25% globally), loneliness, bereavement, ageism, physical health decline, economic insecurity, and caregiver-perpetrated abuse (Yon et al., 2017). These factors complicate timely diagnosis, treatment, and support, leading to under recognised and underserved mental health conditions in older populations. Untreated mental health conditions in older adults impose significant worldwide costs on health and social care systems, stemming from increased hospitalisations, emergency visits, medication usage, long-term care placement, and caregiver lost productivity (Lamoureux-Lamarche et al., 2022). The global economic burden of older adult mental health disorders, running into hundreds of billions of USD annually, underscores the urgent need for effective

intervention strategies (Alzheimers Disease International, 2024)

Artificial intelligence (AI) is transforming healthcare delivery by enhancing access, personalisation, and efficiency across resource-limited settings. Conversational agents (chatbots) have gained prominence among AI applications, simulating human-like interactions through natural language processing and machine learning algorithms to provide 24/7 health information, symptom checking, therapeutic support, medication reminders, and continuous monitoring (Abd-alrazaq et al., 2019). In mental health specifically, chatbots augment traditional services by delivering cognitive behavioural therapy (CBT) exercises, mood tracking, and supportive conversations that target prevalent conditions like depression, anxiety, and loneliness in older adults, while reducing stigma, offering consistent non-judgmental support, and extending limited health workforce capacity (H. Li et al., 2023).

Digital mental health interventions have emerged as promising strategies for expanding service reach and lowering barriers to care. Among these, AI chatbots—ranging from rule-based scripts and machine-learning models to large language model-powered systems—simulate natural dialogue across text, voice, or multimodal interfaces, with early agents demonstrating feasibility for symptom monitoring, psychoeducation, and low-intensity CBT in pilot and small randomised trials (Abd-alrazaq et al., 2019; Haque & Rubya, 2023).

Recent high-quality meta-analyses of general adult populations report statistically significant moderate effects on depression and psychological distress (Hedges' g ≈0.6–0.7), moderated by therapeutic approach, delivery modality, and mobile integration, although effect sizes remain modest, heterogeneity high, and superiority over active digital or human comparators less consistent (H. Li et al., 2023)

For older adults specifically (≥60 years), chatbots present substantial opportunities for companionship, loneliness mitigation, self-management prompts, and guidance to appropriate services. Scoping reviews identify preliminary evidence from small pilots, such as web-based agents MYLO and ELIZA that reduced problem distress, depression-anxiety-stress in controlled trials (n=112), alongside reported improvements in wellbeing and stress, though these remain limited to short-term studies plagued by usability barriers, trust concerns, and operational difficulties particularly with text-driven mobile interfaces (Casu et al., 2024; Mayor, 2025). Broader syntheses consistently note minimal older adult representation in trials dominated by younger/middle-aged samples, with scarce age-stratified outcomes and no coverage of newer LLM-enabled systems (Abd-alrazaq et al., 2019; H. Li et al., 2023).

Implementation success for older adults hinges on addressing usability, perceived usefulness, trust, privacy concerns, and prior technology experience, amid persistent digital divides by age, income, and education despite substantial rises in smartphone and internet adoption over the past decade (Yu & Chen, 2024). Economic considerations are paramount, as mental disorders generate substantial direct health-care costs alongside indirect burdens from functional decline, institutionalisation, and caregiver strain; systematic reviews confirm many prevention/promotion programs prove cost-effective or cost-saving, yet evaluations of digital/chatbot interventions remain scarce with virtually no reporting of costs, QALYs, or formal cost-effectiveness metrics to guide policy investment amid mounting aged-care pressures (Abd-Alrazaq et al., 2020; Le et al., 2021).

Rationale Gap analysis: These literature strands reveal critical gaps including no existing systematic reviews or meta-analyses focused specifically on AI chatbots for mental health support in older adults across all domains, generations from rule-based to LLM-enabled, and care settings (Abd-alrazaq et al., 2019; Mayor, 2025). Prior reviews aggregate across age groups, marginalise older adults as subgroups, or examine adjacent technologies like companion robots or commercial voice assistants (Casu et al., 2024). No integrated synthesis jointly examines clinical effectiveness and economic implications despite evidence that design/implementation features profoundly influence engagement and outcomes (Le et al., 2021; H. Li et al., 2023). The rapid post-2022 emergence of LLM-based systems lacks comprehensive mapping in older populations, while economic evidence remains underdeveloped with minimal model-ready data for aged-care scale-up decisions (Abd-Alrazaq et al., 2020).

Rationale: This systematic review protocol directly addresses these gaps through in-depth quantitative synthesis of AI chatbot effects on key mental health outcomes (such as depression, anxiety, distress, wellbeing) among older adults, drawing from diverse designs including RCTs, quasi-experimental, pre-post, and single-arm trials to capture both rigorous effect estimates and emerging pilot data (H. Li et al., 2023). Building on established methods from prior conversational agent and economic reviews, it will conduct meta-analysis where feasible alongside narrative and dominance-matrix synthesis for heterogeneous economic findings on costs, resource use, and cost-effectiveness (Le et al., 2021; H. Li et al., 2023). Findings will inform subsequent co-design and economic modeling for AI mental health chatbots in global aged-care contexts, equipping clinicians, policymakers under the Aged Care Data and Digital Strategy 2024–2029, and developers with consolidated evidence for safe, acceptable, equitable deployment.

Condition being studied All mental health statuses are considered eligible, including (a) older adults with formally diagnosed mental disorders, (b) those with elevated or subclinical symptoms identified through screening instruments, and (c) unselected community or residential-care samples, defined as older adults recruited from community or long-term care settings without any requirement for mental health problems at baseline (e.g., general primary-care attendees, residents of aged-care facilities, or community-dwelling seniors participating in health promotion programmes).

METHODS

Search strategy Eight bibliographic databases will be searched systematically: MEDLINE (Ovid), Embase, PsycINFO, CINAHL, Scopus, Web of Science Core Collection, EBSCOhost (for relevant nursing/allied health indices not captured elsewhere), and EconLit. Database-specific search strings will be conducted using controlled vocabulary (e.g., MeSH, Emtree) and free-text terms following three core domains: (1) older adults (e.g., "older adult*", "aged", "elder*", "senior*"); (2) mental health conditions and constructs (e.g., "depression", "anxiety", "distress", "loneliness", "mental health", "wellbeing"); and (3) AI conversational technologies (e.g., "chatbot*", "conversational agent*", "virtual agent*", "dialog* system*", "large language model*"). For EconLit and EBSCO-hosted databases, search terms will emphasise economic evaluation concepts such as "cost-effectiveness", "cost-utility", "cost-benefit", "QALY*", "ICER*", and "economic evaluation" in combination with chatbot-related terms.

The time frame will run from 1 January 2014 (to capture the modern era of mobile and AI-based conversational agents) to the date of the final search, with no restrictions on country or clinical setting; searches will be limited to peer-reviewed original research articles in English.

All search strategies will be refined in consultation with an information specialist, and full search strings for each database will be provided in an appendix of the full paper. To supplement database searches, reference lists of included studies and relevant systematic reviews will be screened, and major trial registries will be checked for completed or ongoing chatbot trials in older adults.

Participant or population Studies will be included if participants are adults with a mean or median age of at least 60 years, or if mixed-age samples are reported in which either ≥50% of participants are aged ≥60 years or data for the ≥60 subgroup are separable. All mental health statuses are considered eligible, including (a) older adults with formally diagnosed mental disorders, (b) those with elevated or subclinical symptoms identified through screening instruments, and (c) unselected community or residential-care samples, defined as older adults recruited from community or long-term care settings without any requirement for mental health problems at baseline (e.g., general primary-care attendees, residents of aged-care facilities, or community-dwelling seniors participating in health promotion programmes).

Intervention Eligible interventions are AI chatbots (conversational agents) that simulate dialogue with users using rule-based scripts, machine-learning methods, or foundation models such as large language models. Chatbots may be delivered via text, voice, or multimodal interfaces within standalone applications, web platforms, messaging services, or embedded systems. To be included, the chatbot must provide mental health support, defined as at least one of: psychoeducation, symptom monitoring, self-management support, low-intensity therapeutic techniques (e.g., CBT-based exercises), or support targeting loneliness, social connectedness, or help-seeking.

Comparator Any comparator will be accepted, including usual care, wait-list, information-only or minimal-intervention controls, alternative digital tools, or other active treatments. Studies without a comparator (single-arm trials or case series) are also eligible and will contribute pre–post change estimates only.

Study designs to be included Only quantitative intervention studies will be included, organised into a four-level design hierarchy: (1) randomised controlled trials; (2) quasi-experimental controlled studies; (3) single group pre-post studies; (4) uncontrolled case series or single arm trials with pre-post quantitative data. Purely qualitative studies, qualitative components of mixed-methods studies, and purely technical or simulation papers without human participants will be excluded.

Eligibility criteria Articles must be full peer-reviewed original research articles published in English from 2014 onwards.

Information sources MEDLINE (Ovid), Embase, PsycINFO, CINAHL, Scopus, Web of Science Core Collection, EBSCOhost (for relevant nursing/allied health indices not captured elsewhere), and EconLit.

Main outcome(s) Primary outcomes are validated quantitative measures of depression, anxiety, psychological distress/stress, positive mental health or wellbeing.

Additional outcome(s) Secondary outcomes mainly include economic outcomes such as intervention costs, resource use, incremental costs, ICERs, QALYs, ROI, and related parameters. We also look at adverse outcomes as an additional outcome.

Data management All records identified through database searches will be imported into Nested Knowledge for de-duplication and workflow management. Two reviewers will independently screen titles and abstracts against the eligibility criteria. An initial calibration exercise on approximately 50 records will be conducted to ensure consistent application of criteria. Full texts of potentially eligible studies will then be retrieved and assessed independently in duplicate; disagreements at either stage will be resolved through discussion and, where necessary, consultation with a third reviewer. Reasons for full-text exclusion will be recorded, and the selection process will be presented in a PRISMA 2020 flow diagram (Page et al., 2021).

Data extraction will use a piloted, standardised template (in Excel or within Nested Knowledge). One reviewer will perform extraction and a second will cross-check all entries. Extracted items will include: bibliographic details, country and setting, sample size, age distribution, mental health status, inclusion criteria, intervention characteristics (chatbot type and generation, modality, therapeutic content, interface, duration, level of human support), comparator details, study design level, and outcome measures with timepoints. Numeric data required to compute effect sizes (group means, standard deviations, change scores, event counts) will be extracted for all relevant outcomes. For economic data, information will be captured on perspective, time horizon, cost categories (e.g., intervention, health-care, social care), valuation methods, currency and price year, and any reported ICERs, QALYs, ROI or dominance statements.

Quality assessment / Risk of bias analysis Risk of bias will be evaluated at the outcome level using design-appropriate tools. Randomised controlled trials will be appraised using the Cochrane Risk of Bias 2 (RoB 2) tool, covering randomisation processes, deviations from intended interventions, missing outcome data, outcome measurement, and selection of reported results (Higgins & Cochrane Collaboration, 2019). Non-randomised controlled studies will be assessed with ROBINS-I (Sterne et al., 2016), addressing confounding, selection of participants, classification of interventions, deviations, missing data, outcome measurement, and selective reporting. Single-group pre–post and uncontrolled designs will be assessed using adapted ROBINS-I domains, acknowledging their higher inherent risk of bias. For economic evaluations and cost-effectiveness data, methodological quality will be assessed using the Quality of Health Economic Studies (QHES) instrument, as applied in Le et al.'s review of mental health prevention and promotion interventions (Le et al., 2021). Risk-of-bias and QHES assessments will inform sensitivity analyses and will contribute to GRADE ratings of certainty for key clinical outcomes.

Strategy of data synthesis

Clinical effectiveness and meta-analysis

The primary synthesis will focus on controlled designs (design levels 1–2). Where at least two studies report the same or conceptually similar primary outcome, random-effects meta-analyses will be conducted using Hedges' g standardised mean differences with 95% confidence intervals (J. Li et al., 2025). For each outcome, the post-intervention timepoint closest to the end of treatment will be used in main analyses; longer-term follow-up will be explored in secondary analyses where available. Randomised and quasi-experimental controlled studies will be pooled together, with study design entered as a prespecified moderator in subgroup and meta-regression analyses. Single-group pre–post studies (design level 3) and uncontrolled case series or single-arm trials (design level 4) will be summarised descriptively, with pre–post change statistics presented when available but not formally pooled if data are sparse or highly heterogeneous.

Between-study heterogeneity will be quantified using I^2 and τ^2 statistics, and, where sufficient studies exist, prediction intervals will be reported. Planned subgroup and meta-regression analyses will examine potential effect modifiers, including chatbot generation (rule-based vs machine-learning vs LLM-enabled), modality (text vs voice vs multimodal), setting (community vs primary/specialist care vs residential or long-term care), baseline symptom severity (clinical vs subclinical vs general population), and study design. Publication bias will be investigated using contour-enhanced funnel plots and Egger's test when at least ten studies contribute to a meta-analysis. Certainty of evidence for each primary outcome will be graded using GRADE, considering risk of bias, inconsistency, indirectness, imprecision and publication bias.

Economic impact synthesis and modelling

Economic evidence will be synthesised using a structured multi-step framework adapted from contemporary studies and health-economic review guidance (Gomersall et al., 2015; Le et al., 2021).

Step 1 – Direct pooling and dominance analysis (primary if ≥3 studies)

Where three or more studies report comparable cost-effectiveness information (for example, ICERs expressed as cost per QALY gained from a similar perspective and time horizon), quantitative synthesis will be attempted. Effect measures will be standardised to 2025 Australian dollars using purchasing-power-parity and inflation adjustments, and, where assumptions on comparability are tenable, random-effects meta-analysis of ICERs or net monetary benefit will be undertaken (Chen et al., 2023). Whether or not pooling is feasible, a dominance ranking matrix will be constructed to classify each intervention as: (a) more effective and less costly, (b) more effective and more costly, (c) less effective and less costly, or (d) less effective and more costly than its comparator, providing a transparent summary of value-for-money signals across studies.

Step 2 – De novo Markov modelling (contingent on data availability)

We will attempt a de novo decision-analytic Markov modelling, in which pooled clinical effects (e.g., standardised mean change in depression or anxiety) are mapped to health-state utilities and extrapolated over an appropriate time horizon to estimate ICERs and cost-effectiveness acceptability curves for older adults (Ara & Brazier, 2011; National Institute for Health and Care Excellence, 2013). Implementation of such modelling will, however, depend on the availability of sufficiently homogeneous data to parameterise: (a) baseline transition probabilities between health states (e.g., remission, mild, moderate, severe depression, death), (b) the impact of chatbot interventions on those transitions, (c) utility weights linked to the symptom measures used, and (d) state- and intervention-specific costs in older adult populations. If outcome heterogeneity, lack of valid mapping functions, absence of credible transition data, or sparse cost and utility reporting preclude robust parameterisation, the review will explicitly report that full Markov modelling could not be credibly undertaken.

Step 3 – Enhanced narrative and threshold analysis (fallback / likely minimum output)
If de novo modelling is not feasible, an enhanced narrative economic synthesis will be conducted instead. This will integrate the dominance matrix from Step 1 with simple threshold analyses that explore combinations of plausible effect sizes and per-user costs that would be compatible with conventional willingness-to-pay thresholds for mental health interventions (e.g., cost per QALY gained). This staged approach will still allow clear articulation of likely value-for-money ranges, highlight where AI mental health chatbots may be promising from an economic perspective, and identify key evidence gaps that need to be addressed in future, model-ready trials and evaluations.

Subgroup analysis

We will examine potential effect modifiers to explain heterogeneity ($I^2 > 50\%$) in meta-analyses of primary clinical outcomes, where ≥4 studies per subgroup are available. Prespecified subgroups include: (1) chatbot generation (rule-based vs. machine learning vs. LLM-enabled); (2) delivery modality (text vs. voice vs. multimodal); (3) care setting (community vs. primary/specialist care vs. residential/long-term care); (4) baseline symptom severity (clinical diagnosis vs. subclinical/elevated symptoms vs. unselected/general population); and (5) study design (RCTs vs. quasi-experimental). Tests for subgroup differences (e.g., χ^2) will be reported alongside 95% confidence intervals for between-subgroup contrasts.
Meta-regression will complement subgroups where ≥10 studies permit, modelling continuous moderators like intervention duration, human support level, or mean participant age.

Sensitivity analysis Sensitivity analyses will be conducted to assess the robustness of meta-analytic findings to key methodological decisions and potential biases. For primary clinical outcomes, these will include re-analysis excluding studies judged at high risk of bias (via RoB 2 or ROBINS-I), restricting to randomised controlled trials only, excluding studies with high attrition (>20%), and using alternative effect measures or fixed-effect models where random-effects are primary. Results will be compared to main analyses to determine if conclusions remain consistent. For economic syntheses, sensitivity analyses will test ICER stability by varying key parameters such as unit costs (±20%), time horizons, discount rates (3-5%), and utility mappings, alongside one-way and probabilistic scenarios where data permit. Threshold analyses will explore cost-effectiveness acceptability across willingness-to-pay ranges (e.g., AUD 40,000-70,000 per QALY). These analyses will inform GRADE certainty ratings and highlight influential assumptions.Pls see above.

Language restriction English.

Country(ies) involved Australia (Flinders University, University of Sydney), Ireland (Maynooth University).

Other relevant information The research team is concurrently conducting a parallel systematic review (protocol registered; Sultana et al., 2025; INPLASY Protocol: 8436) that maps the scope, diversity, and characteristics of AI chatbot studies involving older adults. Whereas the parallel review emphasises descriptive mapping across qualitative, quantitative, and mixed-methods designs, the current review employs broader search terms and databases but adopts a more focused scope on intervention studies reporting quantitative clinical outcomes and, where available, economic data. This enables formal meta-analysis and structured economic synthesis to guide policy and implementation. All review stages—from question formulation and rationale, through search strategy and screening, to synthesis and reporting—are undertaken independently de novo.

Keywords artificial intelligence; non communicable diseases; chatbot; older adults; aged care; mental health.

Dissemination plans Findings from this review will be disseminated through presentations at relevant national and international conferences, alongside preparation of manuscripts for submission to peer-reviewed journals in mental health, gerontology, health technology assessment, and digital health fields. Knowledge translation activities will target policymakers, aged care providers, and digital health developers through policy briefs, infographics, and targeted webinars to support evidence-based implementation of AI chatbots for older adults' mental health. Results will be presetned in conferences and papers will be written for submission to peer reviewed journals.

Contributions of each author
Author 1 - Madhan Balasubramanian - Design; Protocol; Search strategy; Screening; Data extraction; Synthesis; Writing first draft.
Email: madhan.balasubramanian@flinders.edu.au
Author 2 - Sharmin Sultana - Design; Search Strategy; Screening; Data extraction; Synthesis.
Email: sharmin.sultana@flinders.edu.au
Author 3 - Sharif Rasel - Design; Protocol.
Email: sharif.rasel@flinders.edu.au
Author 4 - Angie Shafei - Design; Protocol.
Email: angie.abdelshafei@flinders.edu.au
Author 5 - Mahalakshmi Shivashankar - Design; Protocol; Screening.
Email: mahalakshmi.shivashankar@flinders.edu.au
Author 6 - Karen Reynolds - Design; Protocol.
Email: karen.reynolds@flinders.edu.au
Author 7 - Martin Curley - Design; Protocol.
Email: martin.curley@mu.ie
Author 8 - Timothy Schultz - Design; Protocol; Search strategy; Screening; Data extraction; Synthesis; Economic analysis.
Email: timothy.schultz@flinders.edu.au

References

Abd-alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., & Househ, M. (2019). An overview of the features of chatbots in mental health: A scoping review. International Journal of Medical Informatics, 132, 103978. https://doi.org/10.1016/j.ijmedinf.2019.103978

Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020). Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis. Journal of Medical Internet Research, 22(7), e16021. https://doi.org/10.2196/16021

Alzheimers Diseas International. (2024). World Alzheimer Report 2023: Reducing dementia risk: Never too early, never too late. https://www.alzint.org/resource/world-alzheimer-report-2023/

Amir-Behghadami, M., & Janati, A. (2020). Population, Intervention, Comparison, Outcomes and Study (PICOS) design as a framework to formulate eligibility criteria in systematic reviews. Emergency Medicine Journal, 37(6), 387–387. https://doi.org/10.1136/emermed-2020-209567

Ara, R., & Brazier, J. E. (2011). Using Health State Utility Values from the General Population to Approximate Baselines in Decision Analytic Models when Condition-Specific Data are Not Available. Value in Health, 14(4), 539–545. https://doi.org/10.1016/j.jval.2010.10.029

Casu, M., Battiato, S., Guarnera, L., Caponnetto, P., & Triscari, S. (2024). AI Chatbots for Mental Health: A Scoping Review of Effectiveness, Feasibility, and Applications. Applied Sciences, 14. https://doi.org/10.3390/app14135889

Chen, W., Wong, N. C. B., Wang, Y., Zemlyanska, Y., Butani, D., Virabhak, S., Matchar, D. B., Prapinvanich, T., & Teerawattananon, Y. (2023). Mapping the value for money of precision medicine: A systematic literature review and meta-analysis. *Frontiers in Public Health*, *11*, 1151504. https://doi.org/10.3389/fpubh.2023.1151504

Gomersall, J. S., Jadotte, Y. T., Xue, Y., Lockwood, S., Riddle, D., & Preda, A. (2015). Conducting systematic reviews of economic evaluations. International Journal of Evidence-Based Healthcare, 13(3), 170–178. https://doi.org/10.1097/XEB.0000000000000063

Haque, M. D. R., & Rubya, S. (2023). An Overview of Chatbot-Based Mobile Mental Health Apps: Insights From App Description and User Reviews. JMIR mHealth and uHealth, 11, e44838. https://doi.org/10.2196/44838

Higgins, J. P. T. & Cochrane Collaboration (Eds.). (2019). Cochrane handbook for systematic reviews of interventions (Second edition). Wiley-Blackwell.

Lamoureux-Lamarche, C., Berbiche, D., & Vasiliadis, H.-M. (2022). Health care system and patient costs associated with receipt of minimally adequate treatment for depression and anxiety disorders in older adults. BMC Psychiatry, 22(1), 175. https://doi.org/10.1186/s12888-022-03759-9

Le, L. K.-D., Esturas, A. C., Mihalopoulos, C., Chiotelis, O., Bucholc, J., Chatterton, M. L., & Engel, L. (2021). Cost-effectiveness evidence of mental health prevention and promotion interventions: A systematic review of economic evaluations. PLOS Medicine, 18(5), e1003606. https://doi.org/10.1371/journal.pmed.1003606

Li, H., Zhang, R., Lee, Y.-C., Kraut, R. E., & Mohr, D. C. (2023). Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. Npj Digital Medicine, 6(1), 236. https://doi.org/10.1038/s41746-023-00979-5

Li, J., Li, Y., Hu, Y., Ma, D. C. F., Mei, X., Chan, E. A., & Yorke, J. (2025). Chatbot-Delivered Interventions for Improving Mental Health Among Young People: A Systematic Review and Meta-Analysis. Worldviews on Evidence-Based Nursing, 22(4), e70059. https://doi.org/10.1111/wvn.70059

Mayor, E. (2025). Chatbots and mental health: A scoping review of reviews. Current Psychology, 44. https://doi.org/10.1007/s12144-025-08094-2

National Institute for Health and Care Excellence. (2013). Guide to the Methods of Technology Appraisal 2013. National Institute for Health and Care Excellence (NICE). http://www.ncbi.nlm.nih.gov/books/NBK395867/

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, n71. https://doi.org/10.1136/bmj.n71

Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A.-W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., … Higgins, J. P. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. BMJ, i4919. https://doi.org/10.1136/bmj.i4919

Sultana, S., Rasel, S., O'Keefe, V., & Balasubramanian, M. (2025). Artificial intelligence chatbots for supporting mental health in older adults: A mixed-methods systematic review protocol. INPLASY - International Platform of Registered Systematic Review and Meta-analysis Protocols. https://doi.org/10.37766/inplasy2025.10.0110

WHO. (2024). Mental health of older adults [Fact sheet]. https://www.who.int/newsroom/ fact-sheets/detail/mental-health-of-older-adults

Yon, Y., Mikton, C. R., Gassoumis, Z. D., & Wilber, K. H. (2017). Elder abuse prevalence in community settings: A systematic review and meta-analysis. *The Lancet Global Health, 5*(2), e147–e156.
https://doi.org/10.1016/S2214-109X(17)30006-2

Yu, S., & Chen, T. (2024). Understanding older adults' acceptance of chatbots in healthcare delivery: An extended UTAUT model. *Frontiers in Public Health, 12*, 1435329.
https://doi.org/10.3389/fpubh.2024.1435329

Zhong, W., Luo, J., & Zhang, H. (2024). The therapeutic effectiveness of artificial intelligence-based chatbots in alleviation of depressive and anxiety symptoms: A systematic review and meta-analysis. *Journal of Affective Disorders, 352*, 340–352.
https://doi.org/10.1016/j.jad.2024.04.057