


Conversational artificial intelligence interventions to support smoking cessation: A systematic review and meta-analysis

DIGITAL HEALTH
Volume 9: 1–13
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076231211634
journals.sagepub.com/home/dhj



Hollie Bendotti^{1,2,3} , Sheleigh Lawler^{3,4}, Gary C K Chan^{3,5}, Coral Gartner³, David Ireland² and Henry M Marshall^{1,3,6}

Abstract

Background: Conversational artificial intelligence (chatbots and dialogue systems) is an emerging tool for tobacco cessation that has the potential to emulate personalised human support and increase engagement. We aimed to determine the effect of conversational artificial intelligence interventions with or without standard tobacco cessation interventions on tobacco cessation outcomes among adults who smoke, compared to no intervention, placebo intervention or an active comparator.

Methods: A comprehensive search of six databases was completed in June 2022. Eligible studies included randomised controlled trials published since 2005. The primary outcome was sustained tobacco abstinence, self-reported and/or biochemically validated, for at least 6 months. Secondary outcomes included point-prevalence abstinence and sustained abstinence of less than 6 months. Two authors independently extracted data on cessation outcomes and completed the risk of bias assessment. Random effects meta-analysis was conducted.

Results: From 819 studies, five randomised controlled trials met inclusion criteria (combined sample size $n = 58,796$). All studies differed in setting, methodology, intervention, participants and end-points. Interventions included chatbots embedded in multi- and single-component smartphone apps ($n = 3$), a social media-based ($n = 1$) chatbot, and an internet-based avatar ($n = 1$). Random effects meta-analysis of three studies found participants in the conversational artificial intelligence enhanced intervention were significantly more likely to quit smoking at 6-month follow-up compared to control group participants (RR = 1.29, 95% CI (1.13, 1.46), $p < 0.001$). Loss to follow up was generally high. Risk of bias was high overall.

Conclusion: We found limited but promising evidence on the effectiveness of conversational artificial intelligence interventions for tobacco cessation. Although all studies found benefits from conversational artificial intelligence interventions, results should be interpreted with caution due to high heterogeneity. Given the rapid evolution and potential of artificial intelligence interventions, further well-designed randomised controlled trials following standardised reporting guidelines are warranted in this emerging area.

Keywords

Digital health, artificial intelligence, smoking, behaviour change, systematic reviews

Submission date: 16 April 2023; Acceptance date: 3 October 2023

¹Faculty of Medicine, Thoracic Research Centre, The University of Queensland, Chermide, Queensland, Australia

²The Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, Herston, Queensland, Australia

³NHMRC Centre of Research Excellence on Achieving the Tobacco Endgame, School of Public Health, Faculty of Medicine, The University of Queensland, Herston, Queensland, Australia

⁴School of Public Health, Faculty of Medicine, The University of Queensland, Herston, Queensland, Australia

⁵National Centre for Youth Substance Use Research, The University of Queensland, Saint Lucia, Queensland, Australia

⁶The Prince Charles Hospital, Metro North Hospital and Health Service, Chermide, Queensland, Australia

Corresponding author:

Hollie Bendotti, The Prince Charles Hospital, L5 Clinical Sciences Building, Rode Rd, Chermide, QLD 4032, Australia.

Email: h.bendotti@uq.edu.au



Introduction

Tobacco use persists as a leading cause of premature death and contributor to the global burden of disease.^{1,2} Prevalence of tobacco use is disproportionately high in low- and middle-income countries (>80%) compared to high-income countries, thus contributing to the widening of global health burden inequalities.³ Acceptable, accessible, and effective smoking cessation support is vital given the significant short- and long-term individual health benefits. A multifaceted approach to smoking cessation is considered best practice due to the complex interaction of physiological and behavioural factors that influence tobacco dependence,⁴ as well as the relapsing-remitting nature of the condition.⁵ A well-established evidence base exists for the effectiveness of various pharmacotherapies to aid smoking cessation,^{6–9} and their effect is increased when used in combination with behavioural support.¹⁰ Yet these professional interventions, while effective, are limited in their reach^{11,12} and accessibility,^{13,14} and as such many people quit or attempt to quit unassisted despite the lower success of unassisted quitting.^{15,16}

Digital platforms have the capability to expand access to behavioural support for tobacco cessation. Current digital cessation interventions have advanced from computer-generated tailored letters,¹⁷ to internet-based resources,¹⁸ text-messaging programs,¹⁹ and smartphone applications.²⁰ However, these digital platforms in isolation are limited in their ability to provide highly personalised support. According to Mohr's 'Model of Supportive Accountability', engagement with digital health interventions is promoted with the addition of human support by fostering a sense of personal accountability to a competent, trustworthy, and caring coach.²¹ To address this gap between professional tailored support and scalable digital interventions, innovations in artificial intelligence (AI) are being increasingly explored to emulate human support.

Conversational AI, typically referred to as a chatbot or dialogue system, enables two-way communication with users via text and/or audio without human input by employing natural language processing and machine learning algorithms. Modern conversational AI systems commonly use a combination of two approaches. A rule-based approach relies on rules defined and programmed by expert knowledge (e.g., a conversation tree)²² which allows for greater control over the conversation pathway and AI-generated responses. In contrast, a probabilistic approach learns how to classify answers and potentially how to generate responses through exposure to training text.²² Advances in so-called large language models have shown this technique scales with very large training text that gives impressive predictive performance. There are, however, frequent issues with erroneous text being produced that appear legitimate but entirely fictional. There is still considerable work

required to determining the ethical and legal liabilities caused by these fictional responses. Regardless of the technique, through continual use, interactions with AI systems may become more tailored and fluid.²³ Furthermore, the large amount of novel data generated through these systems has the potential to identify patterns in individual and population behaviours, which may enhance our understanding of smoking populations and improve the delivery and outcomes of highly personalised smoking cessation advice.

The use and effectiveness of conversational AI agents in health care is an emerging area of research. Specific health applications of this technology include but are not limited to, the delivery of cognitive behavioural therapy for mental health conditions²⁴ and substance use disorders,²⁵ supporting people with dementia,²⁶ triage support,²⁷ and screening for health conditions.²⁸ A systematic review of 31 studies of AI conversational agents in health care found overall positive or mixed effectiveness, as well as high usability and satisfaction.²⁹ The application of conversational AI for smoking cessation is still in its infancy but holds great potential in expanding tailored support. Benefits may include greater accessibility and may overcome some barriers to seeking counselling support due to fears about potentially negative and stigmatising interactions, which are often experienced by people who smoke, including when interacting with health professionals.^{30,31} A 2022 scoping review reported some benefits of chatbots on smoking cessation but noted variable outcome measures and methodological issues.³² More recently, a systematic review and meta-analysis of conversational agents for smoking cessation found positive effects on cessation outcomes across 6 studies as well as overall high acceptability.³³ However, this analysis included automated synchronous text messaging systems which provided little detail regarding AI paradigms and described constraints on natural language input, and smoking cessation outcomes timepoints included in the meta-analysis varied. Additionally, the relationship between frequency of engagement and cessation outcomes (i.e., a dose-response) was not evaluated. Given the rapid development of conversational AI interventions and limited evidence syntheses, this systematic review and meta-analysis aimed to evaluate the effectiveness of conversational AI interventions and the relationship between level of engagement on tobacco cessation outcomes.

Methods

A study protocol was not registered due to an oversight in handover during project coordinator staffing changes but is attached as Supplemental File 1. This review was completed as part of a project for the World Health Organization (WHO) and the protocol was reviewed and approved by a WHO advisor.

Review questions

1. What effect do conversational AI interventions have on individual tobacco cessation outcomes compared to: (a) No intervention; (b) usual care; or (c) active comparator intervention/s when delivered alone or in combination?
2. What is the relationship between frequency and duration of interaction (i.e., dose) with conversational AI interventions and tobacco cessation outcomes?

Search strategy

A systematic search strategy was developed in collaboration between authors HB, HM and a specialist medical librarian. Databases searched on 3rd June 2022 included PubMed, Embase, CINAHL Complete, APA PsycINFO, CENTRAL, Web of Science. The search was supplemented by searches of four trial registries on 6th June 2022 (ClinicalTrials.gov, WHO ICTRP, EU Clinical Trials Register, and the Australian Clinical Trials Register), and by hand-searching reference lists of included papers and other relevant reviews. MeSH terms included ‘Tobacco Use Cessation’, ‘Smoking Cessation’, ‘Smoking/prevention and control’, ‘Smoking/therapy’, and ‘Artificial Intelligence’. Other AI search terms included ‘chatbot’, ‘conversation* agent’, ‘dialogue system’, ‘virtual agent’, ‘virtual assistant’, ‘machine learning’, ‘natural language’, ‘predictive analysis’, and ‘ELIZA effect’. The full search strategies are attached as Supplemental File 2.

Inclusion/exclusion criteria

Randomised controlled trials (RCTs) that tested the effectiveness of a conversational AI intervention as a smoking cessation aid (either alone or in combination with other behavioural or pharmacological smoking cessation treatments) among people aged 15 years and over who smoked combustible tobacco upon study enrolment (i.e., seeking and/or consented to receive smoking cessation treatment) were included. We included studies that involved a conversational AI tobacco cessation intervention compared with a control group, including no intervention, usual care, and active comparators. Active comparators could include but were not limited to, non-AI tobacco cessation smartphone or tablet applications, internet-based interventions, text-messaging-based interventions, pharmacological cessation aids, telephone counselling, behavioural counselling delivered face-to-face or through non-smartphone apps. Studies with more than one intervention arm/comparator were also included.

AI interventions could be delivered through a smartphone, mobile phone, tablet, computer or webpage, but needed to allow for bidirectional (conversational) communication. Automated messaging (AM) interventions were excluded if the underlying dialogue system had not been

described in detail or did not, or was unlikely to have, employed AI algorithms or allowed for unconstrained natural language input.

Evidence selection

The search and retrieval process is illustrated in Figure 1. Search results were initially imported into Endnote referencing software where duplicates were removed. The updated library was then imported into Covidence where individual studies’ title and abstract were independently screened by one of two reviewers (HB & HM). Full-text screening was independently completed by HB, HM and SL, with any conflicts resolved through discussion between the reviewers or with another senior author (CG). Where required, original study investigators were contacted via email for further information regarding the type of AI system/algorithm to determine inclusion/exclusion. Final determination of AI systems was reached by consensus between an experienced computer scientist who specialises in AI chatbot development (DI) and all other co-authors.

Data extraction

Data were extracted independently and agreed upon by two reviewers (HB and HM) using a form created in Covidence software. Extracted data included data on trial design and characteristics, participant characteristics, intervention and comparator, outcome measures, loss to follow-up, narrative/qualitative findings (if applicable), and authors’ conclusions and declarations. Discrepancies unable to be resolved between the two reviewers were referred to a third reviewer (SL).

Risk of bias appraisal

Quality assessment of included individual studies was completed using Version 2 of the Cochrane risk of bias tool for randomised trials (RoB 2).³⁴ The tool provides an overall risk of bias (low, some concerns, high) after assessment across five domains: (a) Bias arising from the randomisation process; (b) bias due to deviations from intended interventions; (c) bias due to missing outcome data; (d) bias in measurement of the outcome; and (e) bias in selection of the reported result. Appraisal of each study was completed by one study team member (HB) for review by HM, and determination was reached by consensus.

Data synthesis and meta-analysis

Results of included studies were summarised by narrative synthesis to describe the effectiveness of conversational AI interventions, and the relationship between frequency and duration of engagement on cessation outcomes. Random effects meta-analysis of intention-to-treat (ITT)

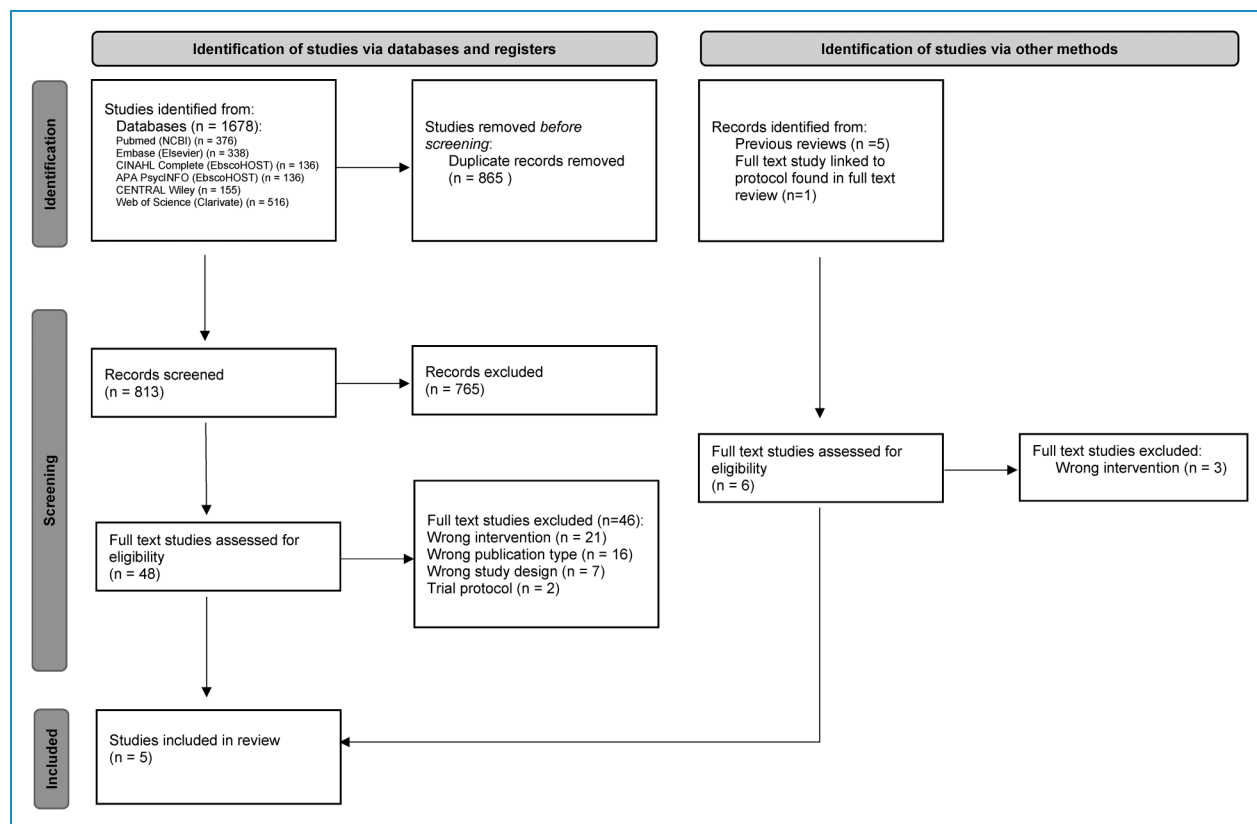


Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram.

results was conducted to estimate the effect of conversational AI interventions on smoking abstinence at 6-month follow-up. This outcome deviated slightly from our protocol outcome of continuous abstinence of at least 6-months due to differences between study designs, but consistency in follow-up timepoints. Three studies were deemed similar enough to be included in this analysis. Abstinence rates were collected and reported as risk ratios. To test if the estimates were sensitive to the inclusion of any single study, leave-1-out analysis was conducted. To test the potential impact of publication bias, trim-and-fill analysis was conducted. All analyses were conducted in R with the metafor package.

Results

Among the 819 unique records identified through database searches ($n=813$) and snowballing ($n=6$), five articles were included in the analysis (Figure 1). Two articles^{35,36} were included after full-text review, and three additional articles were identified from a recent scoping review,^{37,38} and from a trial protocol³⁹ found in the initial search. Extracted data are summarised in Table 1 and Supplemental File 3. Study investigators of one included paper were contacted via email for further information regarding RCT details, however, we received no response

and could therefore only analyse the limited published information.

Participant and study characteristics

Five RCTs conducted in Spain,³⁹ Cyprus,³⁷ Hong Kong,³⁶ Japan³⁸ and online,³⁵ were published between 2018 and 2022 (Table 1). The total combined randomised sample size across the studies was 58,796 participants (range 84–57,214). The minimum age of inclusion for all studies was 18 years. Mean ages of participants and percentage of females ranged from 22 years to 50.7 years, and 24% to 65%, respectively. Participant loss to follow-up ranged from 14% to 80.1% for intervention groups, and 12.2% to 90.2% for comparison groups. Baseline measures of tobacco dependence were reported in four studies and included cigarettes per day (CPD),^{35,37–39} heaviness of smoking index (HSI),³⁹ Fagerstrom Test for Nicotine Dependence (FTND),^{37,38} and time to first cigarette.³⁵ Mean levels of nicotine dependence ranged from low to moderate across studies reporting FTND (range 2.91–5.3) and HSI (range 2.65–2.71) measures. Significant differences in baseline CPD and time to first cigarette were reported in one study but adjusted for in sensitivity analyses.³⁵

Conversational AI interventions included chatbots embedded in a multicomponent smoking cessation app,^{35,38} a chatbot

Table 1. Study characteristics and findings.

Author, year country	Study design and size	Participant characteristics	Intervention	Comparator	Outcome data
Perski 2019 Online (no country/ies specified)	Parallel RCT (automated) <i>Allocated</i> Intervention: 5339 Control: 51 875 <i>LTFU</i> Intervention: 4278 (80.1%) Control: 46825 (90.2%)	Smoke daily or non-daily <i>CPD, mean (SD)</i> Intervention: 16 (11.4) Control: 14.7 (8.9) <i>Time to first cigarette <5mins</i> Intervention: 17.2% Control: 19.3% <i>Setting</i> Online <i>Recruitment</i> Via commercial App Store. People who purchased the “Pro” version of the Smoke Free app.	Smoke Free Pro app with chatbot <i>Intervention Dose</i> 2x daily check in for 4 weeks <i>AI system details</i> AI-driven chatbot	Smoke Free Pro app only	<i>Self-reported continuous abstinence at 4-week follow-up</i> Intervention: 844/5339 (15.8%) Control: 3704/51,875 (7.1%) OR 2.44 (2.25–2.64, $p < .001$) <i>Frequency of engagement with app, median (IQR)</i> Intervention: 16 (65.5) Control: 5 (22) IRR 2.07 (1.97–2.17, $p < .001$)
Olano-Espinosa 2022 Spain	Pragmatic, multicenter RCT <i>Allocated</i> Intervention: 242 Control: 271 <i>LTFU</i> Intervention: 138 (57.0%) Control: 143 (52.8%)	Smoke consecutively (≥ 1 cigarette in past month) <i>Age, mean (SD)</i> Intervention: 49.01 (11.22) Control: 50.66 (10.42) <i>Female</i> Intervention: 146 (60.3%) Control: 158 (58.3%) <i>HSI, mean (SD)</i> Intervention: 2.71 (1.59) Control: 2.65 (1.68) <i>Setting</i> Online <i>Recruitment</i> Via Family practitioners and nurses.	Chatbot delivered intervention (based on 5As: behavioural \pm pharmacological treatment) <i>Intervention Dose</i> Daily interactive contacts, but patient could contact chatbot at any time and decide on frequency and duration of contact. <i>AI system details</i> Probabilistic expert system based on Bayes’ theorem	Standard face-to-face intervention (based on 5As: behavioural \pm pharmacological treatment)	<i>Biochemically validated continuous abstinence at 6 months</i> Intervention: 63/242 (26.0%) Control: 51/271 (18.8%) OR 1.50, 95% CI 1.00–2.31; $P = .05$. <i>6 month abstinence: Intensive contact (>4 contacts lasting >30 min throughout 6 month period) vs non-intensive</i> Intervention Intensive: 68.6% Non-intensive: 40.9% $P = .02$ Control Intensive: 47.6% Non-intensive: 35.4% $P = .30$
Karekla 2020 Cyprus	Pilot Parallel RCT <i>Allocated</i> Intervention:	Smoke daily <i>Age, mean (years)</i> Intervention: 22.5	Internet-based, digital avatar led ACT program.	Wait-list control group	<i>Self-reported 7-day PPA at post-treatment</i>

(continued)

Table 1. Continued.

Author, year country	Study design and size	Participant characteristics	Intervention	Comparator	Outcome data
	49 Control: 35 <i>LTFU</i> Intervention: 22 (44.8%) Control: 7 (20%)	Control group: 22 <i>Female</i> Intervention: 64% Control: 65% <i>FTND, mean (SD)</i> Intervention: 3.19 (2.09) Control: 2.91 (2.02) <i>Setting</i> Three universities <i>Recruitment</i> Flyers posted in university cafeterias and classroom announcements.	<i>Intervention Dose</i> 6 ACT smoking cessation sessions (~25 min each). Frequency of contact spaced out over 3–30 days (i.e. time to complete the entire program and post-assessment varied). <i>AI system details</i> Rule-based. Different continuation of dialogue based on user response.	(questionnaires only)	<i>time-point (after 6 sessions completed)</i> Intervention: 14/49 (28.57%) Control: 4/35 (11.43%) OR 3.10, 95% CI 0.92–10.41, $p=$ 0.067; Mean treatment completion: 44 days
Wang 2018 Hong Kong	Parallel RCT <i>Allocated</i> Intervention: 205 Control: 196 <i>LTFU</i> Intervention: 71 (34.6%) Control: 43 (21.9%)	Past 7-day smoking <i>Age, mean (years)</i> Intervention: 32.8 Control: 33.1 <i>Female</i> Intervention: 78 (39.8%) Control: 84 (41%) <i>Nicotine dependence</i> Not reported <i>Setting</i> Not reported <i>Recruitment</i> Not reported	WeChat groups with conversational agent. <i>AI system details</i> Rule based, context aware dialogue management with natural language processing module.	Smoking cessation information and tips without social support or interactions with other participants	<i>Verified abstinence at 6-month post-intervention follow-up^a</i> Intervention: 27/ 205 (13.2%) Control: 19/196 (9.7%) <i>Participation in conversations vs smoking outcomes</i> Mean number of WeChat posts amongst bioverified quitters was 22.8 compared to 9.1 for participants who did not have bioverified quitting. Participants who had not smoked in the past week were involved in a significantly higher number of conversations than participants who had smoked ($p=0.02$).

(continued)

Table 1. Continued.

Author, year country	Study design and size	Participant characteristics	Intervention	Comparator	Outcome data
Masaki 2020 Japan	Parallel, multi-center, randomised, controlled, open-label trial. <i>Allocated</i> Intervention: 285 Control: 287 <i>LTFU</i> Intervention: 40 (14%) Control: 42 (14.6%)	Currently smoke (smoking history of pack-years ≥ 10) <i>Age, mean (SD)</i> Intervention: 47 (11) Control: 45 (12) <i>Female</i> Intervention: 69 (24.2%) Control: 77 (36.7%) <i>FTND, mean (SD)</i> Intervention: 5.2 (2.0) Control: 5.3 (2.1) <i>Setting</i> 31 smoking cessation clinics <i>Recruitment</i> Via physicians.	CASC system app (24 weeks) +12 week standard smoking cessation treatment (concurrent) The CASC system included interactive counselling with a personalised chatbot and mobile CO checker. Standard smoking cessation treatment: 5 on-site examinations and in-person counselling by a primary physician at each outpatient clinic within 3 months. <i>AI system details</i> Personalised automated guidance system	Control app (24 weeks) +12 week standard smoking cessation treatment (concurrent). Control app provided basic functions of CASC app, not including a chatbot and not accompanied by a mobile CO checker	<i>Biochemically verified continuous abstinence from weeks 9–24</i> Intervention: 182/285 (63.9%) Control: 145/287 (50.5%) OR 1.73, 95% CI 1.24–2.42, $p = 0.001$

RCT: randomised controlled trial; LTFU: loss to follow-up; CI: confidence interval; SD: standard deviation; OR: odds ratio; IRR: incidence rate ratio; HSI: heaviness of smoking index; IQR: interquartile range; CPD: cigarettes per day; FTND: Fagerström test for nicotine dependence; ACT: acceptance and commitment therapy; PPA: point prevalence abstinence; CASC: CureApp smoking cessation; AI: artificial intelligence.

^aMetric and measurement method not reported. Intention to treat results calculated by reviewers.

smoking cessation app,³⁹ a conversational agent embedded in social media groups,³⁶ and an internet-based avatar.³⁷ Three studies compared conversational AI interventions to standard/usual care including in-person visits,³⁹ questionnaires,³⁷ and smoking cessation information without social/professional support.³⁶ Two studies used a control/standard version of the intervention app without the chatbot component.^{35,38} Standard smoking cessation treatment (behavioural and pharmacological) was delivered to both intervention and comparator groups in two studies,^{38,39} with the behavioural component delivered by chatbot in the intervention group in one study.³⁹

Smoking cessation outcomes were biochemically verified in two studies using exhaled carbon monoxide test (threshold < 10 ppm),^{38,39} self-reported in two studies,^{35,37} and one study was unclear as to how “verified abstinence” was ascertained.³⁶ Primary smoking cessation outcomes varied, including self-reported continuous abstinence at 4-week follow-up,³⁵ self-reported 7-day point-prevalence abstinence (PPA) at the posttreatment timepoint,³⁷ biochemically verified continuous abstinence from weeks 9–24,³⁸ and biochemically verified continuous abstinence at 6 months.³⁹ The metric of measuring abstinence at

6 months was not described in the remaining study.³⁶ Secondary smoking cessation outcomes included weekly³⁶ 7-day PPA and at 4 weeks, 8 weeks, 12 weeks,³⁸ 6 months,³⁷ 12 months,³⁸ and biochemically verified continuous abstinence from weeks 9 to 52 and weeks 9 to 12.³⁸ Additional outcomes included changes in quality of life,³⁹ cigarettes smoked per day,³⁷ nicotine dependence,³⁷ intention to quit smoking,³⁷ self-efficacy,³⁷ withdrawal symptoms, cravings, and misperceptions of smoking,³⁸ and time to first lapse after the quit date.³⁸

Effectiveness as a smoking cessation aid

Study results are presented in Table 1. While Wang et al.³⁶ reported per-protocol results only, the data reported in the paper also allowed for an ITT analysis. Of the studies evaluating chatbot smoking cessation apps, only one compared the intervention to face-to-face support. Olano-Espinosa et al.³⁹ found that participants who received behavioural treatment through a chatbot app with pharmacotherapy were significantly more likely to remain abstinent at 6 months compared to those who received standard

behavioural and pharmacological treatment from a doctor or nurse (26.0% vs 18.8%; OR 1.50, 95% CI 1.00–2.31, $P=0.05$). This result lost statistical significance, but the estimate remained relatively unchanged after adjustment for baseline CO-oximetry and bupropion intake (OR 1.52, 95% CI 0.99–2.33, $p=0.53$), and 6-month abstinence was correlated with receiving the chatbot intervention (OR 1.52, 95% CI 0.99–2.33; $p=0.053$) and bupropion prescription (OR 2.81, 95% CI 1.49–5.32; $p=0.001$).³⁹ Similar results were reported by Masaki et al. when comparing the combination of standard in-person counselling alongside either a multicomponent app (CASC) which includes a personalised chatbot, or a control app with basic functions.³⁸ Those who received the CASC intervention were significantly more likely to remain abstinent from weeks 9 to 24 of the trial (63.9% vs 50.5%; OR 1.73 95% CI 1.24–2.42, $p<0.001$) and achieve 7-day PPA at week 24 (72.3% vs 58.2%; OR 1.88, 95% CI 1.33–2.68, $p<0.001$), with results remaining significant at week 52.³⁸

Perski et al. compared two different versions of the same smoking cessation app, with the intervention version containing a proactive and on-demand chatbot.³⁵ When restricting analysis to participants with an elected quit date window within two days before or 14 days after downloading the app (not a pre-specified primary outcome), participants were significantly more likely to report being abstinent after 4 weeks (15.8% vs 7.1%; OR 2.44 (2.25–2.64, $p<0.001$)).³⁵ When the analysis was broadened to include participants without a quit date or a quit date outside this window, the ITT analysis outcomes remained positive and significant (ORadj 1.60, 95%CI 1.51–1.69, $p<0.001$) but per-protocol results were null (ORadj 1.02, 95%CI 0.92–1.13, $p<0.71$).³⁵

Karekla et al.³⁷ investigated the effect of an internet-based digital avatar among young adults (university students). Students were randomised to the Avatar-delivered Acceptance and Commitment Therapy (ACT) behavioural support program or to the waitlist-control group who received questionnaires only.³⁷ Those in the intervention group were more likely to report 7-day PPA at the post-treatment timepoint, although this was not significant (28.6% vs 11.4%; OR 3.10, 95% CI 0.92–10.41, $p=0.067$) which was likely due to the small sample size.³⁷ In the per-protocol analysis the intervention group were 6 times more likely to report abstinence and this was statistically significant (51.9% vs 14.3%; OR 6.46, 95% CI 1.76–23.71, $p=0.005$).³⁷

Finally, Wang et al.³⁶ briefly reported findings comparing a conversational agent embedded in WeChat Groups to smoking cessation information/tips only without social/professional support. The authors stated that details on the RCT were deliberately withheld. In the reported per protocol analysis, a higher proportion of those in the intervention group was verified abstinent at 6 month follow-up (27/134, 20.1%) compared to the control group (19/153, 12.4%),

however, this difference decreased when using an ITT approach (27/205, 13.2% vs 19/196, 9.7%).³⁶ Effect estimates were not provided for this outcome.

Meta-analysis of smoking cessation outcomes

The random effects meta-analysis (Figure 2) of abstinence outcomes at 6-month follow-up ($n=3$) found that participants in the conversational AI intervention were significantly more likely to quit smoking compared to participants in control conditions (RR=1.29, 95% CI (1.13, 1.46), $p<0.001$). Overall, there was no significant heterogeneity ($Q(2)=0.29$, $p=0.867$, $I^2=0\%$), yet this should be interpreted with caution given the small number of studies. The test of funnel plot symmetry did not provide any evidence of asymmetry ($p=0.646$) (Supplemental File 4). Trim-and-fill suggested that there was one potential study missing due to publication bias. Imputing effect size from this potential missing study did not substantially change the effect size (RR=1.26, 95% (1.12, 1.42), $p<0.001$). Leave-1-out sensitivity analysis demonstrated that the estimated effect size may be sensitive to excluding a study. Excluding Masaki, et al. increased the effect size slightly (RR=1.38, 95% CI (1.04, 1.82)). All effect sizes remain highly statistically significant (p ranged from <0.001 to 0.026)

Frequency and duration of engagement

Outcomes related to frequency of engagement or usage rate were reported in four studies,^{35,36,38,39} Masaki et al.³⁸ did not specifically report on interactions with the chatbot component within a multicomponent app, but found that daily use of app functions (e.g., chatbot, diary, video tutorials) was higher among participants who had quit compared to those who had not. Olano-Espinosa et al.³⁹ reported significantly higher abstinence rates among those in the intervention group who interacted with the chatbot intensively (>4 contacts lasting >30 min throughout the 6-month period) compared to those who did not (68.6% vs 40.9%, $p=0.02$). A higher quit rate was also observed in the control group among those who received intensive contact with health professionals compared with those who received less intense support, but the effect was not statistically significant (47.6% vs 35.4% respectively, $p=0.30$).³⁹ Wang et al.³⁶ also noted a significant relationship between greater participation in conversational agent enhanced social-media conversations and smoking abstinence ($p=0.02$). Perski et al.³⁵ found the frequency of engagement with the smoking cessation app was significantly higher for the chatbot version compared to the standard app version, with and without a quit date criterion applied. However, this result was not analysed in relation to smoking cessation outcomes.

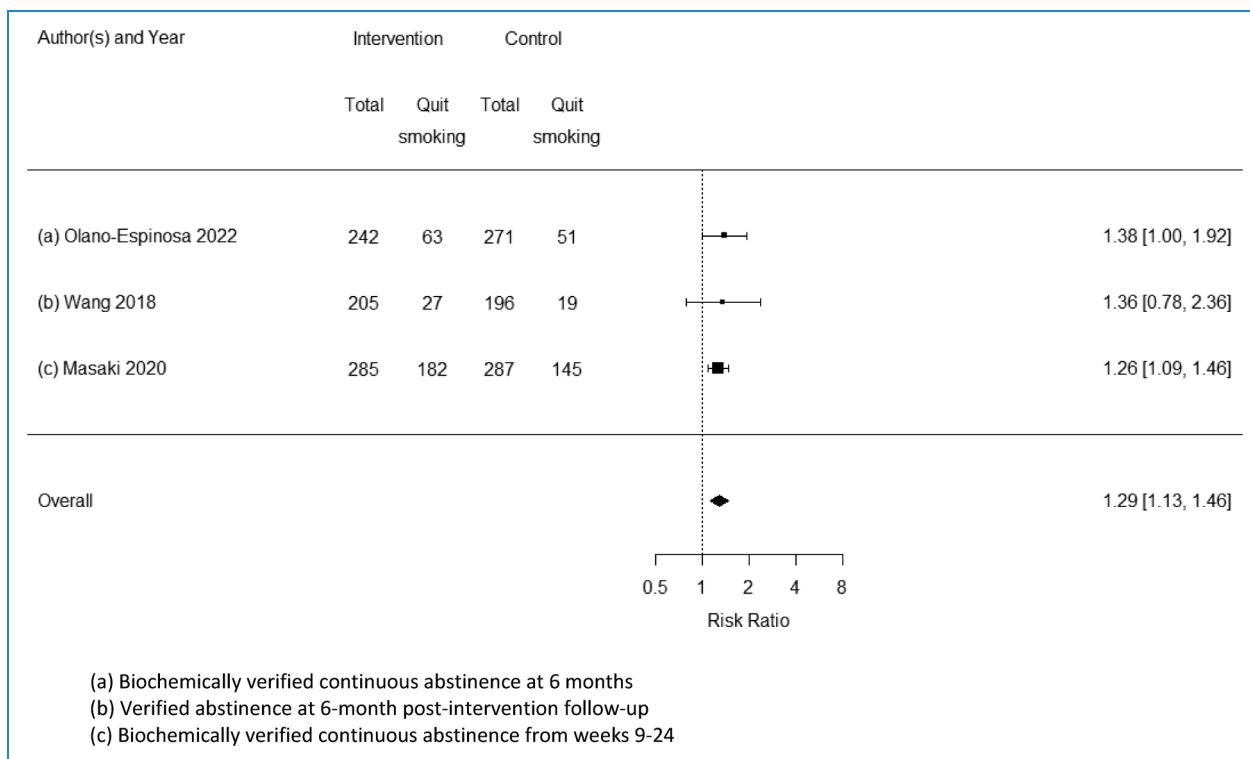


Figure 2. Forest plot of 6-month follow-up abstinence outcomes.

Adverse events

One study reported findings regarding treatment emergent adverse events (AEs). AEs were reported by 64.7% and 63.8% of intervention and control group participants, respectively, and related to the study pharmacotherapy or nicotine withdrawal.³⁹ No specific AEs were related to the intervention app.³⁹

Quality of the evidence

Overall, the risk of bias was high across the studies (Figure 3). As mentioned previously, one study purposively withheld RCT methodological details³⁶ and was therefore deemed a high risk of bias across all domains. For the remaining studies, randomisation was adequate in three studies,^{37–39} while some concerns were raised in the remaining study due to baseline differences between groups because of errors in randomisation coding.³⁵ Low risk of bias was consistent across the four studies for deviations from intended interventions and measurement of outcome data.^{35,37–39} Two of the four studies were deemed high risk of bias due to missing outcome data as there was no evidence of appropriate sensitivity analyses³⁷ or statistical significance was lost when assessing results on a complete-case basis.³⁹ Finally, high risk of bias was found in two studies for selection of the reported result due to

inconsistencies in outcome measurements between trial registry information and published results,³⁷ and, in the second, for multiple analyses allowing for selection of a significant result for a criteria defined outcome not outlined in the research questions.³⁵

Discussion

This systematic review and meta-analysis sought to evaluate the effect of conversational AI interventions on tobacco cessation outcomes, and the relationship between frequency of engagement with these interventions on these outcomes. Our analysis of three studies demonstrated significant positive effects of conversational AI interventions for smoking cessation, however, these results should be interpreted with caution due to the limited number of studies and slight variation in abstinence outcomes. These findings are consistent with a previous systematic review of conversational agents for smoking cessation ($n=6$) which reported favourable abstinence outcomes compared to control groups (OR 1.66, 95% CI=1.33–2.07, $p < 0.001$), but also highlighted variation in study design and quality.³³ Generalisability of results may also be limited given that available measures of nicotine dependence ranged from low to moderate across included studies, warranting further research in cohorts with high levels of nicotine dependence. Additionally, there is limited but

<u>Study ID</u>	<u>D1</u>	<u>D2</u>	<u>D3</u>	<u>D4</u>	<u>D5</u>	<u>Overall</u>	
Perski 2019	!	+	+	+	-	-	+
Wang 2018	-	-	-	-	-	-	!
Karekla 2020	+	+	-	+	-	-	-
Olano-Espinosa 2022	+	+	-	+	+	-	
Masaki 2020	+	+	+	+	+	+	
							D1 Randomisation process
							D2 Deviations from the intended interventions
							D3 Missing outcome data
							D4 Measurement of the outcome
							D5 Selection of the reported result

Figure 3. Risk of bias (RoB2 tool) quality assessment.

promising evidence supporting the ability of conversational AI to increase engagement with mobile health (mHealth) interventions, including a ‘dose-response’ relationship which warrants further comprehensive investigation.

Loss to follow-up (LTFU) was considerable in most studies,^{35–37,39} and greater than 20% in all but one included study. LTFU was markedly higher when an automated RCT approach was used³⁵ compared to trials that involved some level of contact with the study team or clinicians.^{38,39} In contrast to the included mHealth studies, a Cochrane review of 52 studies of combined pharmacotherapy and behavioural interventions found most trials ($n=48$) reported a LTFU rate of less than 20%.⁴⁰ Maintaining participant engagement and motivation is often a challenge for automated mHealth trials, especially for longer duration trials⁴¹ which are required for appropriate evaluation of the effectiveness of tobacco cessation interventions. Regular or semi-regular contact with study team members or clinicians through intervention delivery or follow-up (i.e., a hybrid approach⁴¹) as implemented by Masaki et al.³⁸ and Olano-Espinosa et al.³⁹ may help to improve retention rates within mHealth trials. However, for trials of this nature, the advantage of high retention rates must be weighed against the advantages of large and generalisable samples with minimal human resource requirements achievable through automated trials, where appropriate sensitivity analyses can account for high LTFU rates.

A key challenge in determining inclusion of studies in this review was defining “artificial intelligence”. The definition of AI is broad, often subjective, and differs between experts in the field to such a degree that there is no agreed definition of AI.^{42,43} Our selected studies slightly contrasted with a recent systematic review that assessed the effectiveness of conversational agents for smoking

cessation³³ as we excluded AM systems. While AM systems can be incorporated into conversational AI systems, AM itself does not inherently require or employ AI. We only evaluated bidirectional conversational AI systems, as opposed to solely AM, through consultation and consensus with a senior computer scientist; an approach to eligibility screening which has been employed in previous systematic reviews of AI in health care.⁴⁴ Key terms beyond AI, such as “chatbot”, “rule-based”, “natural language processing”, “probabilistic”, were assumed to be relevant for this review when considered alongside the intervention context, platform and description. Additionally, the true conversational ability of AM systems was assessed and we excluded those that were limited to only one-word utterances by users. Nevertheless, the lack of a clear, objective definition of AI may pose problems in future reviews of efficacy as this technology for smoking cessation further advances.

We observed a difference in how AI intervention trials were reported in health- versus computer science-based journals. Studies published in health-based journals often lacked technical information regarding the AI intervention,^{35,37,38} whereas the study published in a computer science journal lacked detailed trial information and health outcome data necessary to evaluate efficacy and quality.³⁶ While this may be indicative of differing aims and priorities between the fields in terms of research communication, standardised reporting of trials is critical to understanding the potential and efficacy of digital health interventions. A lack of standardised reporting was also noted in a previous systematic review of conversational agents for smoking cessation.³³ The CONSORT-AI extension was developed via expert and stakeholder consultation in 2019, to provide minimum reporting guidelines for trials

evaluating interventions with an AI component.⁴⁵ It adds 14 unique items to the existing CONSORT 2010 reporting guidelines⁴⁶ including, but not limited to, clear descriptions of the AI intervention, how the intervention handles inputs and outputs, and the interaction between humans and AI.⁴⁵ Future trials of AI enhanced smoking cessation interventions, and AI health interventions more broadly, should ensure these guidelines are considered during the planning phase, and followed when reporting results to strengthen future evidence syntheses and individual quality assessments.

Trials of digital health interventions are also not immune from perceived or actual conflicts of interest. Author declarations from three included studies stated connections to the parent company (with or without financial interest) or intellectual property ownership rights to the intervention.^{35,38,39} While the former is well established as a common conflict of interest (COI), previous qualitative research of trial investigators found considerable variability in what was considered and reported as a COI, and highlighted the difficulty in detecting and reporting intellectual COIs.⁴⁷ Trial researchers who have been involved in the design and development of novel or established digital interventions should declare their intellectual COIs given the potential for bias. Furthermore, independent design and evaluation of digital health trials must be ensured and managed in the same way as pharmaceutical trials. While there was some evidence of strategies to improve impartiality in data analysis by using researchers without any IP rights to the intervention,³⁹ the majority of the included studies did not describe management strategies to minimise the influence of potential COIs.

Finally, while early trial results are encouraging, the value of AI-enhanced interventions for health behaviour change must be balanced by acknowledging and addressing potential limitations. Important ethical and legal considerations of AI include: Trust and transparency; data use and privacy; patient safety; bias and health equity; cybersecurity; research and development; governance, testing and evaluation; scope; content decisions; licensing; and third-party involvement.⁴⁸ Conversational AI may prove to be an important tool in overcoming equity issues by providing scalable personalised smoking cessation support in which individual access/engagement is not limited by felt stigma. However, we must be cognizant of the risk of digital health interventions increasing disparities between those who do and do not have the skills and access to them, including geographic location, given that many social determinants of health (e.g., healthcare, education, community and social support) rely on digital access and literacies.⁴⁹ Given the disparities in smoking prevalence between demographic groups and variation in tobacco control legislation and smoking cessation services between countries, digital inclusion of priority populations on a national and international scale is vital for the success

of AI interventions for smoking cessation. While we cannot surmise this potential impact from the current data, given that smoking cessation is multifaceted, we foresee the role of conversational AI is to complement current standards of care (i.e., pharmacotherapy and behavioural support) not to replace them entirely. Therefore, a lack of smoking cessation services, or health services more broadly, due to geographical location could influence the uptake and/or effectiveness of this technology. Future programs of research should seek to promote digital inclusion by; understanding user needs and ensuring the intervention is of benefit, low cost and importance to users; building trust, and ensuring transparency and security of personal data use and storage; multidisciplinary collaboration; and evaluating wider outcomes.⁵⁰

Conclusion

Conversational AI for tobacco cessation is a rapidly evolving field of research. At present, there is limited but promising evidence supporting the use of conversational AI to assist tobacco cessation. The ability of the technology to increase engagement with an intervention may be a key driver for positive cessation outcomes. A high level of heterogeneity between all studies was identified. Future trials should seek to employ more standardised methods and measures to RCT designs, alongside standardised reporting guidelines, to improve our ability to assess conversational AI intervention effectiveness, study quality and the AI systems used.

Acknowledgements: The authors would like to thank The Prince Charles Hospital Librarian, Jana Waldmann, for her assistance in developing the search strategy and completing the databases searches. We would also like to acknowledge colleagues from the World Health Organization (WHO) for their feedback on the study protocol.

Contributorship: HB and HM conceived the study. HB, HM, SL and CG were involved in protocol development. HB, HM, SL, CG and DI contributed to data collection and extraction. GC and HB completed data analysis. HB wrote the first draft of the manuscript, and all co-authors reviewed and provided edits to the manuscript. All authors approved the final version of the manuscript.

Declaration of conflicting interests: The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: This study was completed for a project commissioned by the WHO, but no funding or additional resources were allocated to this individual review. HB has received a New Investigator grant from The Prince Charles Hospital Foundation (NI2021-31).

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by an Investigator Grant (GNT1178331) from the National Health and Medical Research Council (NHMRC) awarded to HM. HB receives a top-up scholarship from the Commonwealth Scientific and Industrial Research Organisation. CG receives funding from NHMRC grants (GNT1198301, GNT2019252) and is supported by an ARC Future Fellowship (FT220100186).

Guarantors: HB & HM.

ORCID ID: Hollie Bendotti  <https://orcid.org/0000-0001-5078-4809>

Supplemental material: Supplemental material for this article is available online.

References

- World Health Organisation. *WHO report on the global tobacco epidemic, 2017: Monitoring tobacco use and prevention policies*. Geneva: World Health Organisation, 2017. https://www.who.int/tobacco/global_report/2017/en/
- Stanaway JD, Afshin A, Gakidou E, et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018; 392: 1923–1994.
- World Health Organisation. *World health statistics 2020: Monitoring health for the SDGs, sustainable development goals*. Geneva: World Health Organisation, 2020.
- Benowitz NL. Nicotine addiction. *N Engl J Med* 2010; 362: 2295–2303. <https://apps.who.int/iris/handle/10665/332070>
- Steinberg MB, Schmelzer AC, Richardson DL, et al. The case for treating tobacco dependence as a chronic disease. *Ann Intern Med* 2008; 148: 554–556.
- Lindson N, Chepkin SC, Ye W, et al. Different doses, durations and modes of delivery of nicotine replacement therapy for smoking cessation. *Cochrane Database Syst Rev* 2019. DOI: 10.1002/14651858.CD013308.
- Howes S, Hartmann-Boyce J, Livingstone-Banks J, et al. Antidepressants for smoking cessation. *Cochrane Database Syst Rev* 2020. DOI: 10.1002/14651858.CD000031.pub5.
- Hartmann-Boyce J, Chepkin SC, Ye W, et al. Nicotine replacement therapy versus control for smoking cessation. *Cochrane Database Syst Rev* 2018. DOI: 10.1002/14651858.CD000146.pub5.
- Mishra A, Maiti R, Mishra BR, et al. Comparative efficacy and safety of pharmacological interventions for smoking cessation in healthy adults: A network meta-analysis. *Pharmacol Res* 2021; 166: 105478.
- Hartmann-Boyce J, Hong B, Livingstone-Banks J, et al. Additional behavioural support as an adjunct to pharmacotherapy for smoking cessation. *Cochrane Database Syst Rev* 2019; 6: Cd009670.
- Roddy E, Antoniak M, Britton J, et al. Barriers and motivators to gaining access to smoking cessation services amongst deprived smokers—a qualitative study. *BMC Health Serv Res* 2006; 6: 147.
- van Rossem C, Spigt MG, Kleijns JR, et al. Smoking cessation in primary care: Exploration of barriers and solutions in current daily practice from the perspective of smokers and healthcare professionals. *Eur J Gen Pract* 2015; 21: 1–7.
- Tall JA, Brew BK, Saurman E, et al. Implementing an anti-smoking program in rural-remote communities: challenges and strategies. *Rural Remote Health* 2015; 15: 3516.
- Hutcheson TD, Greiner KA, Ellerbeck EF, et al. Understanding smoking cessation in rural communities. *J Rural Health* 2008; 24: 116–124.
- Smith AL, Chapman S and Dunlop SM. What do we know about unassisted smoking cessation in Australia? A systematic review, 2005–2012. *Tob Control* 2015; 24: 18–27.
- Soulakova JN and Crockett LJ. Unassisted quitting and smoking cessation methods used in the United States: analyses of 2010–2011 tobacco use supplement to the current population survey data. *Nicotine Tob Res* 2017; 20: 30–39.
- Strecher VJ, Kreuter M, Den Boer DJ, et al. The effects of computer-tailored smoking cessation messages in family practice settings. *J Fam Pract* 1994; 39: 262–270.
- Taylor GMJ, Dalili MN, Semwal M, et al. Internet-based interventions for smoking cessation. *Cochrane Database Syst Rev* 2017. DOI: 10.1002/14651858.CD007078.pub5.
- Whittaker R, McRobbie H, Bullen C, et al. Mobile phone-based interventions for smoking cessation. *Cochrane Database Syst Rev* 2016. DOI: 10.1002/14651858.CD006611.pub4.
- Barnett A, Ding H, Hay KE, et al. The effectiveness of smartphone applications to aid smoking cessation: a meta-analysis. *Clinical eHealth* 2020; 3: 69–81.
- Mohr DC, Cuijpers P and Lehman K. Supportive accountability: A model for providing human support to enhance adherence to eHealth interventions. *J Med Internet Res [Internet]* 2011; 13: e30. PMC3221353.
- Almusharraf F, Rose J and Selby P. Engaging unmotivated smokers to move toward quitting: design of motivational interviewing-based chatbot through iterative interactions. *J Med Internet Res* 2020; 22: e20251.
- Stone P, Brooks R, Brynjolfsson E, et al. Artificial intelligence and life in 2030. In: *One hundred year study on artificial intelligence: Report of the 2015-2016 study panel*. Standing Committee of the One Hundred Year Study of Artificial Intelligence. Stanford, CA: Stanford University, September 2016: 6–52. <http://ai100.stanford.edu/2016-report>
- Fitzpatrick KK, Darcy A and 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; 4: e19.
- Prochaska JJ, Vogel EA, Chieng A, et al. A therapeutic relational agent for reducing problematic substance use (woebot): development and usability study. *J Med Internet Res* 2021; 23: e24850.
- Russo A, D’Onofrio G, Gangemi A, et al. Dialogue systems and conversational agents for patients with dementia: the human–robot interaction. *Rejuvenation Res* 2018; 22: 109–120.

27. Chang P, Sheng Y-H, Sang Y-Y, et al. Developing a wireless speech- and touch-based intelligent comprehensive triage support system. *Comput Inform Nurs* 2008; 26. DOI: 10.1097/01.NCN.0000304754.49116.b4.
 28. Håvik R, Wake JD and Flobak E, A Conversational interface for self-screening for ADHD in adults. In: Bodrunova S, et al. *Internet science. INSCI 2018. Lecture notes in computer science*, Vol. 11551. Cham: Springer, 2019; 133–144. DOI: 10.1007/978-3-030-17705-8_12.
 29. Milne-Ives M, de Cock C, Lim E, et al. The effectiveness of artificial intelligence conversational agents in health care: Systematic review. *J Med Internet Res* 2020; 22: e20346.
 30. Wigginton B, Morphett K and Gartner C. Differential access to health care and support? A qualitative analysis of how Australian smokers conceptualise and respond to stigma. *Crit Public Health* 2017; 27: 577–590.
 31. Boland VC, Mattick RP, McRobbie H, et al. I'm not strong enough; I'm not good enough. I can't do this, I'm failing": A qualitative study of low-socioeconomic status smokers' experiences with accessing cessation support and the role for alternative technology-based support. *Int J Equity Health* 2017; 16: 196.
 32. Whittaker R, Dobson R and Garner K. Chatbots for smoking cessation: Scoping review. *J Med Internet Res* 2022; 24: e35556.
 33. He L, Balaji D, Wiers RW, et al. Effectiveness and acceptability of conversational agents for smoking cessation: a systematic review and meta-analysis. *Nicotine Tob Res* 2022: ntac281. DOI: 10.1093/ntr/ntac281.
 34. Sterne JAC, Savovic J, Page MJ, et al. RoB 2: A revised tool for assessing risk of bias in randomised trials *BMJ* 2019; 366: 14898. DOI: 10.1136/bmj.14898.
 35. Perski O, Crane D, Beard E, et al. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. *Digit Health* 2019; 5. DOI: 10.1177/2055207619880676.
 36. Wang H, Zhang Q, Ip M, et al. Social media-based conversational agents for health management and interventions. *Computer (Long Beach Calif)* 2018; 51: 26–33.
 37. Karekla M, Savvides SN and Gloster A. An avatar-led intervention promotes smoking cessation in young adults: A pilot randomized clinical trial. *Ann Behav Med* 2020; 54: 747–760.
 38. Masaki K, Tateno H, Nomura A, et al. A randomized controlled trial of a smoking cessation smartphone application with a carbon monoxide checker. *npj Digit Med* 2020; 3: 35.
 39. Olano-Espinosa E, Avila-Tomas JF, Minue-Lorenzo C, et al. Effectiveness of a conversational chatbot (dejal@bot) for the adult population to quit smoking: Pragmatic, multicenter, controlled, randomized clinical trial in primary care. *JMIR mHealth UHealth* 2022; 10: e34273.
 40. Stead LF, Koilpillai P, Fanshawe TR, et al. Combined pharmacotherapy and behavioural interventions for smoking cessation. *Cochrane Database Syst Rev*. 2016. DOI: 10.1002/14651858.CD008286.pub3.
 41. Kakkar AK, Sarma P and Medhi B. Mhealth technologies in clinical trials: opportunities and challenges. *Indian J Pharmacol* 2018; 50: 105–107.
 42. Wang P. What do you mean by "AI"? *Front Artif Intell Appl* 2008; 171: 362–373. https://cis.temple.edu/~pwang/Publication/AI_Definitions.pdf
 43. Wang P. On defining artificial intelligence. *J Artif Gen Intell* 2019; 10: 1–37.
 44. Plana D, Shung DL, Grimshaw AA, et al. Randomized clinical trials of machine learning interventions in health care: a systematic review. *JAMA Netw Open* 2022; 5: e2233946–e.
 45. Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *The Lancet Digital Health* 2020; 2: e537–ee48.
 46. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 Explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Br Med J* 2010; 340: c869.
 47. Østengaard L, Lundh A, Tjørnhøj-Thomsen T, et al. Influence and management of conflicts of interest in randomised clinical trials: qualitative interview study. *Br Med J* 2020; 371: m3764.
 48. McGreevey JD III, Hanson CW III and Koppel R. Clinical, legal, and ethical aspects of artificial intelligence-assisted conversational agents in health care. *JAMA* 2020; 324: 552–553.
 49. Sieck CJ, Sheon A, Ancker JS, et al. Digital inclusion as a social determinant of health. *npj Digit Med* 2021; 4: 52.
 50. Government Digital Service. A checklist for digital inclusion – if we do these things, we're doing digital inclusion. Government of the United Kingdom. 2014. <https://gds.blog.gov.uk/2014/01/13/a-checklist-for-digital-inclusion-if-we-do-these-things-were-doing-digital-inclusion/#second-identifier>.
-