

Effectiveness and Acceptability of Conversational Agents for Smoking Cessation: A Systematic Review and Meta-analysis

Linwei He MSc¹, Divyaa Balaji MSc², Reinout W. Wiers PhD^{3,4}, Marjolijn L. Antheunis PhD¹, Emiel Krahmer PhD¹

¹Department of Communication and Cognition, Tilburg School of Humanities and Digital Sciences, Tilburg University, Tilburg, The Netherlands

²Amsterdam School for Communication Research, University of Amsterdam, Amsterdam, The Netherlands

³Addiction Development and Psychopathology (ADAPT)-Lab, Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

⁴Centre for Urban Mental Health, University of Amsterdam, Amsterdam, The Netherlands

Corresponding Author: Linwei He, MSc, Department of Communication and Cognition, Tilburg School of Humanities and Digital Sciences, Tilburg University, Warandelaan 2, 5037 AB, Tilburg, The Netherlands. Telephone: 31(0)644911989; E-mail: l.he_1@tilburguniversity.edu

Abstract

Introduction: Conversational agents (CAs; computer programs that use artificial intelligence to simulate a conversation with users through natural language) have evolved considerably in recent years to support healthcare by providing autonomous, interactive, and accessible services, making them potentially useful for supporting smoking cessation. We performed a systematic review and meta-analysis to provide an overarching evaluation of their effectiveness and acceptability to inform future development and adoption.

Aims and Methods: PsycInfo, Web of Science, ACM Digital Library, IEEE Xplore, Medline, EMBASE, Communication and Mass Media Complete, and CINAHL Complete were searched for studies examining the use of CAs for smoking cessation. Data from eligible studies were extracted and used for random-effects meta-analyses.

Results: The search yielded 1245 publications with 13 studies eligible for systematic review (total $N = 8236$) and six studies for random-effects meta-analyses. All studies reported positive effects on cessation-related outcomes. A meta-analysis with randomized controlled trials reporting on abstinence yielded a sample-weighted odds ratio of 1.66 (95% CI = 1.33% to 2.07%, $p < .001$), favoring CAs over comparison groups. A narrative synthesis of all included studies showed overall high acceptability, while some barriers were identified from user feedback. Overall, included studies were diverse in design with mixed quality, and evidence of publication bias was identified. A lack of theoretical foundations was noted, as well as a clear need for relational communication in future designs.

Conclusions: The effectiveness and acceptability of CAs for smoking cessation are promising. However, standardization of reporting and designing of the agents is warranted for a more comprehensive evaluation.

Implications: This is the first systematic review to provide insight into the use of CAs to support smoking cessation. Our findings demonstrated initial promise in the effectiveness and user acceptability of these agents. We also identified a lack of theoretical and methodological limitations to improve future study design and intervention delivery.

Introduction

Cigarette smoking is one of the major causes of preventable death and premature diseases, contributing to more than 6 million deaths per year worldwide.¹ Recent surveys indicate that almost 70% of smokers have the intention to quit smoking, and over half of them have made a quit attempt in the past year; however, the actual cessation rate remains low.^{2,3} Research shows that aided quit attempts are more likely to succeed than unaided quit attempts,⁴ and developing effective cessation interventions has been a public health priority. There is good evidence for the effectiveness of therapist-delivered interventions, such as brief advice, individual and group counseling, and telephone counseling.⁴ While such support provided by healthcare professionals is effective, the use of it remains low. A large-scale survey among smokers in a number of western countries shows that less than 20% of

smokers have made use of cessation services during a quit attempt.⁵ The major reasons included the need for on-site visits and lengthy waiting times due to staff shortages.⁶ To combat these challenges, innovative digital tools such as conversational agents (CAs) have become increasingly popular in the healthcare domain.

CAs are computer programs that use artificial intelligence to simulate a conversation with users through natural language.⁷ Examples of CAs range from digital assistants such as Siri (Apple) and Alexa (Amazon) to customer service agents available on commercial websites.^{8,9} Automated synchronous text-messaging systems are also considered as a form of CA as they allow two-way communication between a human user and the computer system. More recently, CAs are being used to assist healthcare services because they are always accessible, can engage users in human-like conversations, and

Received: August 1, 2022. Revised: November 23, 2022. Accepted: December 5 2022.

© The Author(s) 2022. Published by Oxford University Press on behalf of the Society for Research on Nicotine and Tobacco.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

provide personalized contents to multiple users simultaneously. Evidence has begun to accumulate around the benefits of CAs in diverse fields including disease diagnoses, medication monitoring, mental health, and risk communication during the recent COVID-19 pandemic.^{11–14} However, to the best of our knowledge, there is no systematic investigation of the effectiveness of CAs for smoking cessation. Additionally, the accessibility of CAs may appeal to certain hard-to-reach groups and populations that face physical and time constraint barriers (eg, lack of resources and lengthy waiting time) to accessing traditional interventions, such as at-risk youth and individuals with low socioeconomic status.¹⁰ To extend access to rural, hard-to-reach individuals, it is essential to gain insights into user experience and design CAs that meet their needs and preferences.

Given the potential and the infancy of the use of CAs for smoking cessation, there is a clear need to systematically summarize the available evidence regarding the use of these agents to inform future development and adoption of them. As the effectiveness and user experience are intricately related,¹⁵ the aim of this systematic review and meta-analysis is to evaluate both the effectiveness and acceptability of CAs for smoking cessation. We aim to identify (1) characteristics, functions, and core conversational features of the CAs; (2) theoretical and technical foundations of the CAs; (3) user experience and needs; and (4) limitations and areas for future work. The results of this review will provide insights into the future design and implementation of CAs in smoking cessation interventions.

Methods

This systematic review was conducted in accordance with the protocol (CRD42022313055) registered at the International Prospective Register of Systematic Reviews (PROSPERO) and is reported following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.¹⁶

Eligibility Criteria

Studies were included if they were published in English and if they: (1) had been peer-reviewed; (2) reported on the use of a CA or synonymous system (eg, automated synchronous text-messaging systems) that allowed autonomous two-way interaction without support from a human; (3) addressed smoking cessation or relapse prevention; and (4) reported outcomes evaluated from the direct end-users. Studies were excluded if the communication was one-way where the CA messages could not be responded to by the user (eg, reminders and pop-up notifications). Studies that only discussed the design or the development of the agents and reported no evaluation or the evaluation was not from end-users (eg, protocols and cost-effectiveness studies) were also excluded.

Considering the infancy of this field, we did not apply any restriction regarding the year of publication, population groups, or geographical locations to provide a comprehensive overview of the evolution of CAs for smoking cessation. Study design was not considered as a key restriction in this review. We included both randomized controlled trials (RCTs) and non-randomized studies (eg, before-and-after designs), observational studies, and qualitative studies.

Search Strategy

A systematic search of peer-reviewed literature was performed using the following databases: PsycInfo, Web of Science, ACM Digital Library, IEEE Xplore, Medline, EMBASE,

Communication and Mass Media Complete, and CINAHL Complete, with no restriction on publication date. Reference lists of included studies were searched to identify additional relevant literature. Two sets of search terms were used for the literature search. The first set addressed CAs and included synonyms such as “dialogue system”, “digital agent”, and “virtual coach”. The second set addressed smoking cessation and included keywords such as “tobacco control”, “smok* reduction”, and “smok* abstinence”. Full search strategies can be found in [Supplementary File 1](#). Last searches were completed on March 1, 2022.

Study Selection

One thousand two hundred and forty-five records were identified from database searches and were imported to the review software Rayyan.¹⁷ After removing duplicates, 874 studies were screened based on title and abstract. A pilot screening ($k = 90$) was conducted independently by two reviewers (LH and DB). With substantial agreement (Cohen's $k = 0.71$) achieved, one reviewer (LH) carried out the remaining abstract screening. Full texts of studies considered as potentially eligible were then screened by two reviewers (LH and DB) using the software EndNote, and any discrepancies were resolved by group discussion. Finally, reference lists of included studies were also searched and screened, resulting in 13 studies included in the present review with a substantial inter-coder agreement (Cohen's $k = 0.71$). See [Figure 1](#) for the detailed selection process.

Data Extraction and Risk-of-Bias Assessment

Included studies were coded to capture information regarding the study characteristics (eg, study design, sample type and size, demographics), the CA system (eg, modality, initiator, artificial intelligence techniques, communication channel), the intervention characteristics (eg, theoretical framework, comparator, intervention duration, and interaction frequency), and the outcomes (smoking-related outcomes and user experience outcomes). A standardized Excel form was devised by the author team to record the data. One author (LH) extracted the data from all included papers.

Risk-of-bias assessment of each study was conducted to ascertain the validity and reliability of the study methods and findings to inform the results of this review. The Cochrane Collaboration Risk of Bias tool¹⁸ was used to appraise the quality of RCTs and the Risk of Bias in Non-randomized Studies of Interventions (ROBINS-I) for non-randomized interventions.¹⁹

Data Synthesis and Analysis

A narrative synthesis was conducted to summarize the results of all the included studies and describe the effectiveness and user experience of current CAs for smoking cessation. Additionally, a meta-analysis was conducted with six RCTs that reported on abstinence (both biochemically verified and self-reported) to create an overall effect size. Abstinence rates were extracted and presented as odds ratios. The majority of studies reported abstinence at multiple endpoints (eg, at 1-month and 3-month follow-up), and we included outcomes measured at the longest endpoint unless a primary outcome was specified by the authors. A random effects model was adopted for the meta-analysis given the variation in the study designs and intervention content.²⁰ Statistical heterogeneity was assessed using the I^2 statistic with an I^2 larger than 50% indicating significant heterogeneity and subgroup analysis is then warranted.²¹ Visual

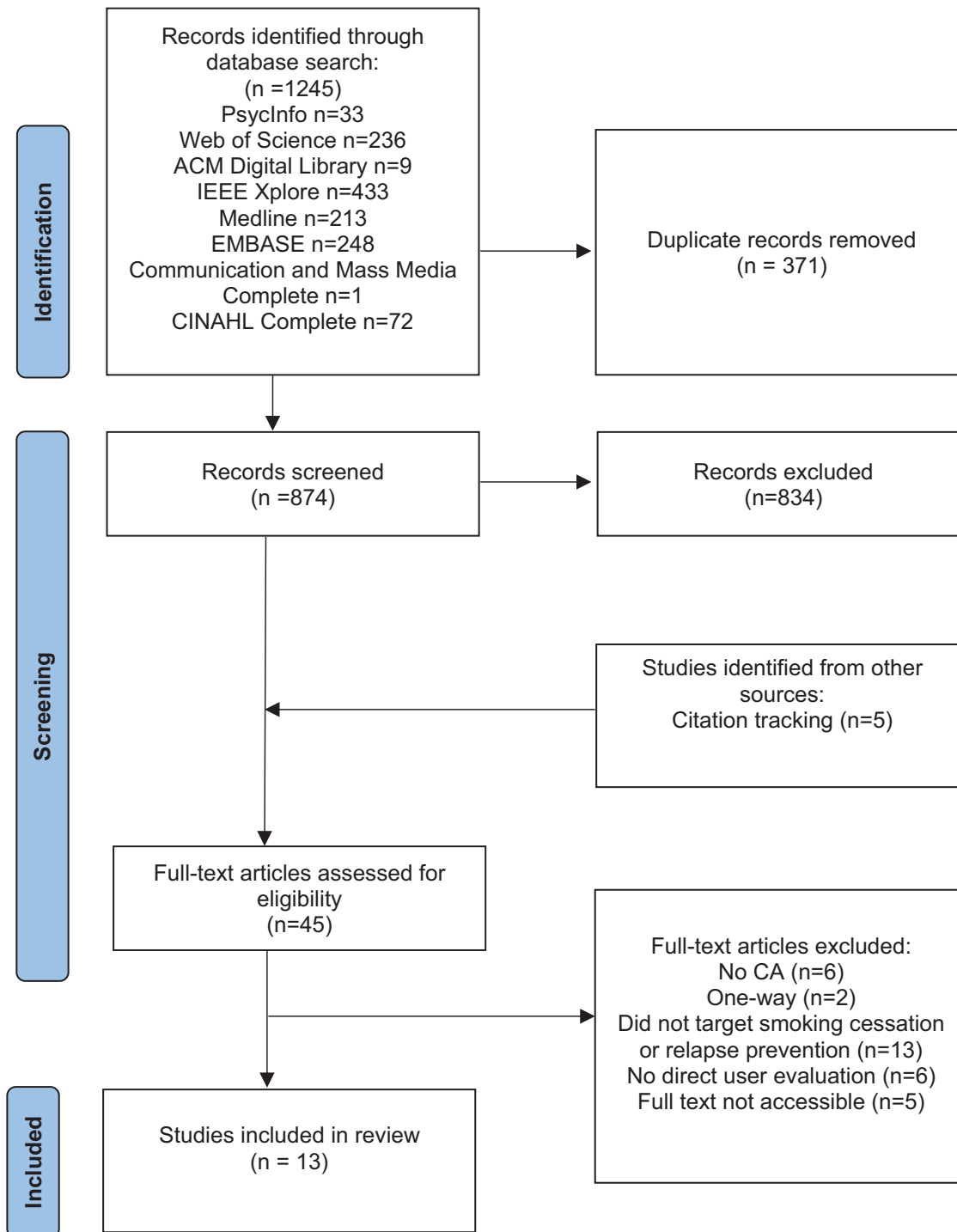


Figure 1. PRISMA chart of the selection process.

inspection of the funnel plot was used to assess publication bias. Additional sensitivity analyses were carried out excluding trials providing only self-reported abstinence and with a high risk of bias. All the statistical analyses were conducted using the Cochrane software ReviewManager (RevMan).²²

Results

Study Characteristics

The key characteristics of the included studies are summarized in Table 1 and more details can be found in Supplementary

File 2. The 13 included studies were conducted between 2012 and 2021 in eight countries including the United States,²³⁻²⁶ the United Kingdom,²⁷ Canada,²⁸ Japan,²⁹⁻³¹ Hong Kong,³² Vietnam,³³ Cambodia,³⁴ and Switzerland.³⁵ There is considerable variation in study design, CA characteristics, and intervention design across the studies. Most studies used mixed methods combining quantitative (eg, experiments) and qualitative (eg, interviews) approaches, and the majority reported on both cessation outcomes and user experience outcomes.^{23-29,32-34} Seven RCTs were identified, and the six remaining included studies utilized either pre-post

Table 1. Characteristics of Included Studies

Authors	Country	Study design	Total N	Sociodemographic	Control	Total intervention duration	Interaction frequency	System platform	Initiator	Primary cessation-related outcome
Abdullah et al. ²³	United States	Pre-post	6	Male (66.7%), mean age 56.3 years, white (50%), at least some college education (67%), unemployed (83%)	NA	14 days	Daily	Tablet: App	System	Self-report 7-day abstinence at post intervention
Abroms et al. ²⁵	United States	RCT	503	Male (34.4%), mean age 35.7 years, white (78.5%), at least some college education (78.1%)	Guidebook	3 months	Daily, gradually decreased to weekly	Phone: SMS	Both	Repeated point prevalence abstinence, saliva verified
Abroms et al. ²⁴	United States	Pre-post	23	Male (56.5%), mean age 20.9 years, white (56.5%)	NA	4 weeks	Daily, gradually decreased to weekly	Phone: SMS	Both	Self-report abstinence at post intervention
Almusharraf et al. ²⁸	Canada	Pre-post	121	Male (48.4%), mean age 35.2 years, paid employment (67.2%)	NA	21.3 minutes	Once	Website	System	Self-reported perceived quitting benefits
Bui et al. ³⁴	Cambodia	RCT	50	Male (100%), mean age 44.2 years, years of formal education 8.5.	App without the CA component	8 weeks	Daily	Phone: App	System	7-day abstinence at 2-month follow-up, CO verified
Calvaresi et al. ³⁵	Switzerland	Pre-post	270	Male (27%)	NA	NI	NI	Phone: App	system	Self-reported abstinence at 3-month follow-up
Jiang et al. ³³	Vietnam	RCT	100	Male (98%), mean age 38.9 years, at least high school education (77%)	Assessment message	6 weeks	Daily	Phone: SMS	Both	7-day abstinence at 12-week follow-up, CO verified
Kato et al. ³¹	Japan	Pre-post	177	Male (63.3%), mean age 44.6 years	NA	24 weeks	Daily	Phone: App	Both	Continuous abstinence 9–12 weeks, salivary verified
Masaki et al. ²⁹	Japan	Pre-post	51	Male (71%), mean age 43.3 years	NA	24 weeks	Daily	Phone: App	Both	Continuous abstinence 9–24 weeks, CO verified
Masaki et al. ³⁰	Japan	RCT	490	Male (75%), mean age 46 years	App without the CA element	24 weeks	Daily	Phone: App	Both	Continuous abstinence 9–24 weeks, CO verified
Perski et al. ²⁷	United Kingdom	RCT	6111	NI	App without the CA element	1 month	Daily	Phone: App	Both	Self-report abstinence at post intervention
Wang et al. ³²	Hong Kong	RCT	134	Male (75%), mean age 46 years	Educational information via phone	2 months	Daily	Phone: App	Both	Self-report 7-day abstinence at post intervention
White et al. ²⁶	United States	RCT	200	Male (22.5%), median age 45 years, predominantly non-Hispanic white (74.5%)	Extra peer mentoring	47–57 days	Daily	Phone: SMS	Both	7-day abstinence at 3-month follow-up, saliva vary verified

CA = conversational agent; RCT = randomized controlled trial; NI = no information; NA = not applicable.

or post-designs. Among the RCTs, the most common comparator was same application or devices as the experimental group but without the active CA element^{27,30,32,34}; other comparators included guidebook²⁵ and message with low intensity for data collection purposes only.³³ Finally, one RCT comparing the CA with an extra active element (ie. CA plus peer mentoring)²⁶ was also included for insights on user experience. Regarding the wide variation in study design, most studies did not mention prespecified or registered study plans (except for²⁶ and ²⁷).

The 13 studies represented a total of 8236 participants who were recruited in both clinical settings^{23,29,30,33,34} and nonclinical settings.^{24-28,31,32} Study sample sizes ranged from 6²³ to 6111.²⁷ For nine of the studies, the mean age of participants fell into the 35–64 years range. Two studies recruited younger participants with a mean age between 20 and 34 years. Two further studies did not provide information on this. For studies that provided socio-demographic information of the participants, most of their samples were balanced in terms of gender, with predominantly white participants with at least some college education. Most of the studies provided information regarding participants' baseline smoking behavior (except^{32,35}), and the mean number of daily cigarette consumption was between 11 and 20. Five studies assessed participants' nicotine dependence level (also referred to as tobacco use disorder, according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [DSM-5]) mainly using FTND (the Fagerström Test for Nicotine Dependence³⁶) and TDS (the Tobacco Dependence Screener³⁷), and the majority were moderately to highly nicotine-dependent smokers.^{24,25,29-31}

Risk of Bias

Of the 13 studies included, two were judged as having a low risk of bias,^{25,33} five having a moderate or unclear risk of bias,^{27,29-31,34} and five had a high risk of bias.^{23,24,26,32,35} The most common reason that contributed to a high risk of bias was possible deviations from the intended intervention, which included unequal amount of intervention received or adhered to by the participants. For example, the time that participants spent on interacting with the CA varied and not all participants read all the texts from or sent messages to the CA, and the on-demand option was not used by all participants.^{24,25,33} Only having self-reported abstinence was another source that frequently contributed to a moderate or high risk of bias.^{23,24,35} See [Supplementary File 3](#) for the detailed assessment for individual studies and a visual summary of the assessed risk of bias.

CA Interventions

The 13 included studies represented 9 unique CA systems. Among them, six were CAs operating in a chatroom format^{23,27-32,35} and three were automated synchronous text-messaging systems.^{24-26,33,34} The CA systems operated on different platforms, with seven accessed on an app on the phone,^{27,29-32,34,35} one on a tablet,²³ one integrated into a website,²⁸ and four using SMS services.^{24-26,33} Only one CA was embodied (ie. with visual representations of the agent).²³ The majority (7 out of 13) of the CAs took natural written language,^{27-32,35} and 4 took responses from participants using fixed keywords.^{24-26,33} The remaining two studies did not specify whether the CA processed natural language.^{23,34} A large proportion of studies (10 out of 13) did not provide

description on the technical architecture of the CA systems, and those who did all used a rule-based infrastructure, and no natural language generation techniques were involved.^{28,32,35} Most of the CA interaction was individualized (ie, the agent interacted with individual users separately), and one study was in a group chat setting where the agent facilitated the interaction between multiple users.³² The majority of the systems allowed both the agent and the user to initiate the interaction^{24-26,29-33} while the rest allowed only the agent to start the conversation. Lastly, two studies^{23,34} provided participants with mobile devices for the intervention, and participants from the rest of the studies used their own devices.

The CAs primarily aimed at smoking cessation promotion, and the dialogs were designed with the implementation of a range of behavioral change theories, including Social Cognitive Theory,^{24,25,33} Transtheoretical Model,^{33,34} Motivational Interviewing,²⁸ and Cognitive Behavioral Theory.³³ The remaining six studies did not provide information on the theoretical bases of CA development.

The CA interventions varied considerably in duration and frequency, ranging from one-time interaction²⁸ to daily interaction over 24 weeks.²⁹⁻³¹ Multi-session interventions all involved daily interaction, with a few studies gradually decreasing the frequency to at least weekly.^{24,25} Six of the studies used the CAs as a stand-alone intervention, while the rest combined the CAs with other intervention components such as personalized web portals and emails^{24,25} and other non-interactive educational information in the app.^{27,29-31,34}

Effectiveness Evaluation

All 13 studies assessed smoking cessation-related outcomes but with great variation in the measures. Eight studies reported biochemically verified abstinence using either salivary cotinine or carbon monoxide readings^{25,26,29-34} while three studies included self-reported abstinence only.^{24,27,35} In addition, two studies did not assess abstinence but reported on other cessation-related outcomes such as a reduction in cigarettes smoked, adopting a smoking ban in the household, and perceived benefits of smoking cessation.^{23,28} Most studies that assessed abstinence (both biochemically verified and self-reported) reported abstinence at multiple endpoints, with the shortest being 2 weeks²⁴ and the longest assessed at 52 weeks follow-up.^{29,30} Except for one study³⁶ that assessed outcomes 6 weeks after the participants had stopped using the CA, all other studies assessed the immediate response directly after use.

A meta-analysis including six RCTs was performed to examine the overall effectiveness of the interventions ($n = 7625$). The study by White et al.²⁶ was excluded since they compared the CA with an extra intervention element, and isolating the effect of CA intervention was therefore not possible. The sample-weighted odds ratio indicated that CA interventions significantly increased the odds of abstinence compared to control groups (odds ratio = 1.66, 95% CI = 1.33% to 2.07%, $p < .001$) (see [Figure 2](#)). Heterogeneity statistics indicated low heterogeneity ($I^2 = 2.5\%$) and no subgroup analysis was performed. Examination of the funnel plot indicated some asymmetry suggesting possible publication bias where unpublished studies with negative or smaller effects might be missing (see [Figure 3](#)).

We performed sensitivity analyses by iteratively removing studies that used only self-reported measures,^{27,32} with a high risk of bias,³² and with a considerably large sample weight.²⁷

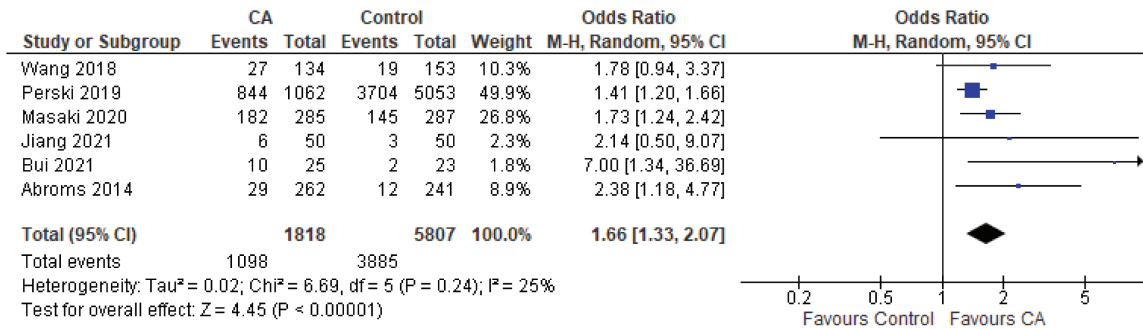


Figure 2. The effectiveness of conversational agent intervention for smoking cessation.

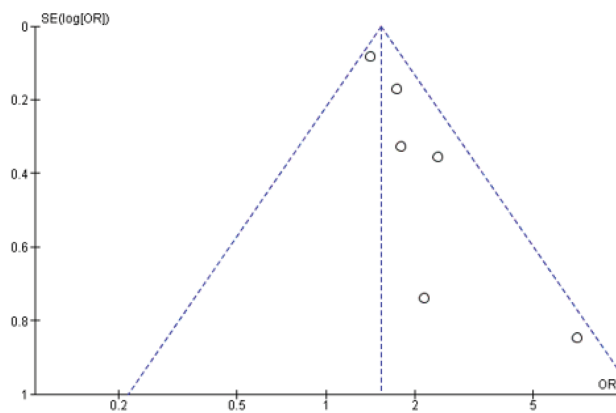


Figure 3. Funnel plot assessing publication bias.

The results can be found in [Supplementary File 4](#). In summary, removing the studies had no impact on the overall effects of the interventions, although heterogeneity slightly increased to 38% when the study with a high risk of bias was removed.

The remaining non-RCTs with pre-post designs also reported positive results regarding the interventions. Four of the non-RCTs reported abstinence with at least a 28.6% abstinence rate at follow-ups, compared to baseline.^{24,29,31,35} Moreover, 100% of participants ($n = 23$) in the study of Abroms et al.²⁴ and 50% of participants ($n = 3$) in the study of Abdullah et al.²³ reported having made at least one quit attempt after the intervention.

Acceptability Evaluation

Generally, from studies that reported user experience outcomes ($k = 10$), participants reported high acceptability of the CA interventions. The majority of studies reported positive user experiences using a number of self-reported metrics. For example, participants reported that they were generally satisfied with the CA interventions,^{23,33} that they liked the agents or the interventions,^{24,26,29} that they felt comfortable interacting with the agents,³⁴ that they found the intervention useful²⁴ and easy to use,³⁴ and that they would recommend the systems to other people.^{24,34} In addition to standard questionnaires, three studies that asked for open-ended user feedback also revealed a few challenges and barriers in using CA.^{23,28,33} The most commonly mentioned barriers included lack of human-likeness, lack of response coherence, limitations in the agents' understanding and responding abilities, and inappropriate responses that frustrated users possibly due to the pre-scripted nature. Some participants in the study by Abdullah et al.²³ mentioned that they would like more self-

paced interaction, for example, being able to start the interaction as needed and not being pushed by the agent to set a quit date. Participants in the study by Jiang et al.³³ expressed a strong interest in a more interactive approach with more freedom in texting in natural language. Last, some participants suggested extending the interaction duration.^{23,33}

Studies reported on engagement with the CA interventions using a number of different metrics, and reporting was inconsistent. Most studies operationalized and measured engagement as the amount of usage, and they found overall high engagement. For example, studies reported that over 85% of participants sent at least one message to the agent,²⁵ that 75% of participants read all messages,²⁴ that participants interacted with the agent for 36 out of 42 days,³³ and that the agents increased the frequency of logins²⁷ and increased the discussion in the chat.³²

Discussion

Principal Findings

CAs are becoming increasingly implemented in health promotion as a promising tool to aid public health services. This review is the first to assess the effectiveness and acceptability of CAs in promoting smoking cessation. The CA interventions reviewed in this study varied considerably regarding system and intervention features, participants characteristics, comparison conditions, and cessation outcomes. Despite the heterogeneity in study designs and outcome evaluation methods used, all studies reported positive effects on smoking cessation-related outcomes, ranging from biochemically verified abstinence to self-reported perceived benefits of cessation. In addition, a meta-analysis of six RCTs demonstrated significantly higher odds of cessation when a CA was involved compared to control groups. This finding provides initial promise for the potential efficacy of CAs in promoting smoking cessation. However, an indication of potential publication bias was observed. Further, there was considerable variation in measurements and reporting of outcomes, which together indicate that the observed positive effects should be interpreted with caution. For example, some studies measured self-reported reduction in cigarette consumption after the intervention,²³ whereas others assessed whether participants made any quit attempts during the intervention.²⁴ One study did not measure cessation directly but asked for self-reported benefits of cessation, to indicate the potential effectiveness of the CA.²⁸ Such inconsistencies in evaluation methods have been observed in previous systematic reviews on the use of CAs in other health-related areas.^{38,39} Notably, the majority of the studies did not preregister their study designs and analysis

plans (except for Refs ²⁶ and ²⁷), which reduces the replicability and reproducibility of the results.⁴⁰ Moreover, among the RCTs, less than half of the studies specifically mentioned the adherence to the intention-to-treat principle or the use of the complete sample set,^{26,30,34} which might lead to overestimation of the actual effect and potentially explained the observed publication bias.⁴¹ In conclusion, the present review found initial promise for the use of CA in assisting smoking cessation, but the actual effects need to be future examined, using more rigorous preregistered research, using standardized outcomes and analysis approaches.

Regarding the user experience of the CAs, results were mixed but a majority reported positive experiences, which illustrates initial user acceptance of the agents as the intervention vehicle. Interestingly, participants valued some aspects of the agents that are usually seen as unique in human–human counseling. For example, participants appreciated that the CA was supportive in interpersonal relationships and that they connected with the CA on a personal level.²³ This is consistent with research demonstrating that people respond in social ways to computers as if they would respond to humans.⁴² Users' appreciation of the relational aspect (as unique in human–human communication) is essential for long-term engagement, which predicts sustained smoking cessation.^{38,43} People often need several quit attempts to achieve abstinence, and long-term support is valued by smokers in the quitting process.^{44,45} However, the majority of the CAs in this review did not specify designs for long-term engagement, and the coincidence finding of users appreciating the relational aspects calls for more attention in future designs to build up the overall positive user experience observed in this review.

There were also a number of other limitations of the agents noted in this review. In addition to the aforementioned lack of relational aspects that hinder long-term engagement, CAs in the included papers and for healthcare oftentimes use pre-scripted utterances and only allow constrained user input to ensure controllability and avoid unwanted harm,^{46,47} while user experience might be compromised in the long term, because of the somewhat predictable nature of scripted interactions. In addition to the limitations in the CAs' technical capabilities, participants expressed frustration when the agents gave them too much pressure to name a quit date²³ or when the agents asked for self-reflection which participants found uncomfortable to answer.²⁸ This echoes the joint research agenda in human–computer interaction and addiction counseling that focuses on the core relational factors such as being empathic and understanding and adapting to patients' motivational states instead of taking a confrontational and demanding approach.^{48,49} Lack of control was another important barrier mentioned by participants, especially in the pace and length of the interaction. Participants prefer interacting with the agents at their own pace and in their own words (ie, more free texts rather than predefined keyword options). This is consistent with previous reviews on the use of CA interventions demonstrating that autonomy is a key factor in the experience with the agents and in ultimate behavioral change.⁷ To summarize, CAs appear to be an overall acceptable format for smoking cessation interventions, with areas for improvement suggested for future development.

Future Directions

Theoretical frameworks for designing the contents of the chatbot intervention are essential to understanding the po-

tential mechanism of behavioral change. However, theoretical frameworks were not reported in many of the studies included in this review. The lack of a theoretical foundation for chatbot intervention development has been noted in previous research.³⁸ Among the reviewed studies that reported on theoretical frameworks, Social Cognitive Theory was the most frequently used framework.^{24,25,33} In addition, we found that the majority of behavioral change techniques were designed to increase persuasion (eg providing contingent rewards and facilitating goal setting),²⁷ while relational strategies received less attention. Relational communicative strategies are being increasingly implemented in recent CA development, especially for mental health,⁴⁸ and resulted in positive effects.⁵⁰ However, current CAs for smoking cessation seem to focus less on the emotional and relational needs of the patients, while the patient–counselor relationship is essential to achieve long-term goals.⁵¹ Nonetheless, two studies in this review cited frameworks that emphasize stages of change,^{28,34} such as the Transtheoretical Model and Motivational Interviewing, which echoes participants' need for empathic, non-demanding, and self-paced interventions that emerged from the qualitative feedback. Future studies could benefit from considering the literature on counseling and consulting and on human–computer interactions to develop CAs that can build and maintain a positive relationship with the users. For example, Bickmore et al.^{48,52} outlined a list of key theoretical considerations in designing relational agents for clinical purposes, such as therapeutic alliance and empathy. It is also encouraged to report on the theoretical foundation and behavioral change techniques to help ascertain the active ingredient of the interventions.

Engagement with the CAs was not reported consistently and the measures varied across studies, limiting the systematic assessment of the interventions. For example, several studies adopted the average length of interaction as an indicator of engagement,^{24,25} while some used frequency of interaction or logins to measure engagement.^{26,27,33} The number of user-initiated interactions was also used to assess user engagement.^{25,29,33} Previous reviews on the use of healthcare artificial intelligence systems also noted the inconsistency in measuring and reporting user engagement.^{7,38} Such inconsistencies could be problematic as it hinders the comparison across studies, while systematic summarization is essential given the infancy of this research area.³⁸ Therefore, standardization of conceptualization and measures of user engagement is needed in future research. For example, Perski et al.⁵³ conceptualized user engagement with digital behavioral change interventions in terms of both subjective experience and objective usage behaviors and suggested both scale questionnaires and standardized usage indicators as viable metrics for user engagement.

Similarly, the technical infrastructure was not reported in the majority of the studies, while technology can have a great impact on how users experience the intervention. With regard to chatbot dialogs, only three studies mentioned the dialogue management system and they were all rule-based.^{28,32,35} The utterances were human-authored and were sent to the users based on pre-defined rules. Rule-based agents are the majority of the current healthcare systems to ensure safety and controllability in sensitive domains.⁵⁴ In terms of user input, most studies allowed a combination of fixed keywords and free texts, even though some participants mentioned that the agents' ability to understand free texts was limited. In

summary, the CAs in the reviewed studies seem to be largely constrained, thus ensuring the controllability and consistency in the delivery of content.⁵⁵ However, they are less able to adapt to individual conversations and are, therefore, perceived as less natural and less engaging, which may hinder long-term interactions with users.^{38,55} Future research is encouraged to extend the initial CA effectiveness observed in this review, by improving the flexibility of the natural language interactions (moving away from pre-scripted interactions), while at the same time balancing this with controllability (making sure that interactions are constrained and do not go off the rails). Furthermore, various modalities (eg, embodied or non-embodied, speech- or text-based) of CA have been used for other health domains such as mental health and physical activities,⁴⁷ while the agents identified in this review were all text-based. It will be important for future studies to compare different modalities and provide further insights into what works and for whom. Technical characteristics can have a great impact on users' perception of and experience with the agents, and ultimately influence the intervention outcomes.^{46,56} Future studies are encouraged to provide descriptions of the technical infrastructure of the agents to allow a better understanding of the relationship between technology and health outcomes.

In summary, there is initial support for the use of CAs in smoking cessation, although no strong conclusions can be drawn yet, given indications of publication bias, lack of preregistered studies, and standardized outcomes. To better realize their potential and to have a more robust evaluation, we call for future research to further examine the long-term effect using more rigorous and standardized approaches.

On a broader note, CAs are being increasingly used in healthcare in general, with many products being developed and implemented over the decade showing positive effects in various behavior domains.^{12,57} Many of the findings reported in this paper generalize to CAs for stimulating other healthy behaviors. While recent advancement has enabled CAs to simulate human-like interactions, we would like to note that they should be considered as a supplementary tool rather than a replacement of human healthcare providers. CAs are useful as an on-demand addition when users need support and healthcare professionals only have limited availability. However, it is not yet clear whether the current generation of CAs lives up to its potential due to the great variation in their designs and evaluation. To ensure that they provide responsible and accountable support, further developments are needed to foster long-term engagement and understand the user-CA interactions. Moreover, it should be noted that most of the reviewed studies were conducted in developed countries and did not investigate sociodemographic factors in people's experience with CAs. Recent rapid advances in technology can unintentionally increase health disparities⁵⁸ by overlooking certain characteristics, needs, or preferences of potential users. To better realize the promise of the CAs and reduce the disparities in the use of such technology, it is crucial that future research include more diverse populations and consider disadvantaged groups in the process of CA development.

Conclusion

Despite the limited body of evidence, this review provides initial promise for the effectiveness and acceptability of CAs for smoking cessation. Interventions using a CA reported better

cessation outcomes compared with control conditions. User experience appeared to be overall positive, while special needs for relational components and self-paced interaction also emerged. The potential of such agents in assisting healthcare resulted in a rapid increase in both industrial development and publications, while standard measures and evaluations of such agents are still lacking, impeding the generalizability of the results. Additionally, theoretical and technical foundations of the agents need more attention in future development. Overall, the present findings demonstrate the potential of CAs for smoking cessation and suggest a clear need for further rigorous design and evaluation of such agents.

Supplementary Material

A Contributorship Form detailing each author's specific involvement with this content, as well as any supplementary data, are available online at <https://academic.oup.com/ntr>.

Funding

This work is supported by the Look Who's Talking Project (official project code: 406.D1.19.054), funded by the Dutch Research Council (NWO). The funder has no role in study design, data collection and analysis, decision to publish or preparation of manuscripts.

Acknowledgments

We would like to thank project members of Look Who's Talking for their valuable insights and suggestions during the discussions we had.

Declaration of Interests

The authors declare that they have no competing interests.

Data Availability

The full search strategy can be found in supplementary materials. The full list of screened articles can be obtained from the corresponding author upon request.

References

1. Reitsma MB, Fullman N, Ng M, *et al*. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990–2015: a systematic analysis from the Global Burden of Disease Study 2015. *Lancet*. 2017;389(10082):1885–1906.
2. Hagerman CJ, Hoffman RK, Vaylay S, Dodge T. Implementation intentions to reduce smoking: a systematic review of the literature. *Nicotine Tob Res*. 2020;23(7):1085–1093.
3. Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting smoking among adults—United States, 2000–2015. *MMWR Morb Mortal Wkly Rep*. 2017;65(52):1457–1464.
4. Stead LF, Koilpillai P, Fanshawe TR, Lancaster T. Combined pharmacotherapy and behavioural interventions for smoking cessation. *Cochrane Database Syst Rev*. 2016;3:CD008286.
5. Gibson JE, Murray RL, Borland R, *et al*. The impact of the United Kingdom's national smoking cessation strategy on quit attempts and use of cessation services: findings from the International Tobacco Control Four Country Survey. *Nicotine Tob Res*. 2010;12(suppl 1):S64–S71.

6. Taylor GM, Dalili MN, Semwal M, *et al.* Internet-based interventions for smoking cessation. *Cochrane Database Syst Rev.* 2017;9(9):CD007078.
7. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. *JMIR Ment Health.* 2019;6(10):e14166.
8. Adam M, Wessel M, Benlian A. AI-based chatbots in customer service and their effects on user compliance. *Electron Markets.* 2021;31(2):427–445.
9. McTear M, Callejas Z, Griol D. *The Conversational Interface: Talking to Smart Devices.* Cham, Switzerland: Springer International Publishing; 2016.
10. Bryant J, Bonevski B, Paul C, McElduff P, Attia J. A systematic review and meta-analysis of the effectiveness of behavioural smoking cessation interventions in selected disadvantaged groups. *Addiction.* 2011;106(9):1568–1585.
11. Miner AS, Laranjo L, Kocaballi AB. Chatbots in the fight against the COVID-19 pandemic. *NPJ Digit Med.* 2020;3(1):1–4.
12. Laranjo L, Dunn AG, Tong HL, *et al.* Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc.* 2018;25(9):1248–1258.
13. Chaix B, Bibault J-E, Pienkowski A, *et al.* When chatbots meet patients: one-year prospective study of conversations between patients with breast cancer and a chatbot. *JMIR Cancer.* 2019;5(1):e12856.
14. He L, Basar E, Wiers RW, Antheunis ML, Krahmer E. Can chatbots help to motivate smoking cessation? A study on the effectiveness of motivational interviewing on engagement and therapeutic alliance. *BMC Public Health.* 2022;22(1):1–14.
15. Smyth A, Syrek C, Reins JA, *et al.* User experience predicts the effectiveness of a gamified recovery app. *Z Gesundh Wiss.* 2018;13(4):319–326.
16. Shamseer L, Moher D, Clarke M, *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ.* 2015;349:g7647–g7647.
17. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev.* 2016;5(1):210. doi:10.1186/s13643-016-0384-4.
18. Higgins JP, Altman DG, Gøtzsche PC, *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ.* 2011;343:d5928.
19. Sterne JA, Hernán MA, Reeves BC, *et al.* ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ.* 2016;355:i4919.
20. Bornstein M, Hedges L, Higgins J, Rothstein H. *Introduction to Meta-analysis.* West Sussex, England: John Wiley & Sons; 2009.
21. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* 2003;327(7414):557–560.
22. *Review Manager (RevMan) [Computer program]. Version 5.4.* Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration; 2020.
23. Abdullah AS, Gaehde S, Bickmore T. A tablet based embodied conversational agent to promote smoking cessation among veterans: a feasibility study. *J Epidemiol Glob Health.* 2018;8(3–4):225–230.
24. Abroms LC, Ahuja M, Kodl Y, *et al.* Text2Quit: results from a pilot test of a personalized, interactive mobile health smoking cessation program. *J Health Commun.* 2012;17(suppl 1):44–53.
25. Abroms LC, Boal AL, Simmens SJ, Mendel JA, Windsor RA. A randomized trial of Text2Quit: a text messaging program for smoking cessation. *Am J Prev Med.* 2014;47(3):242–250.
26. White JS, Toussaert S, Thurl J, *et al.* Peer mentoring and automated text messages for smoking cessation: a randomized pilot trial. *Nicotine Tob Res.* 2020;22(3):371–380.
27. Perski O, Crane D, Beard E, Brown J. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. *Digit Health.* 2019;5:2055207619880676.
28. Almusharraf F, Rose J, Selby P. Engaging unmotivated smokers to move toward quitting: design of motivational interviewing-based chatbot through iterative interactions. *J Med Internet Res.* 2020;22(11):e20251.
29. Masaki K, Tateno H, Kameyama N, *et al.* Impact of a novel smartphone app (CureApp smoking cessation) on nicotine dependence: prospective single-arm interventional pilot study. *JMIR MHealth UHealth.* 2019;7(2):e12694.
30. 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(1):1–7.
31. Kato A, Tanigawa T, Satake K, Nomura A. Efficacy of the assure smoking cessation program: Retrospective study. *JMIR MHealth UHealth.* 2020;8(5):e17270.
32. Wang H, Zhang Q, Ip M, Lau JTF. Social media-based conversational agents for health management and interventions. *Computer.* 2018;51(8):26–33.
33. Jiang N, Nguyen N, Siman N, *et al.* Adaptation and assessment of a text messaging smoking cessation intervention in Vietnam: pilot randomized controlled trial. *JMIR MHealth UHealth.* 2021;9(10):e27478.
34. Bui TC, Sopheab H, Businelle MS, *et al.* Mobile-health intervention for smoking cessation among Cambodian people living with HIV: a mixed-methods pilot study. *AIDS Care.* 2021;34(4):10.
35. Calvaresi D, Calbimonte J-P, Dubosson F, Najjar A, Schumacher M. Social network chatbots for smoking cessation: agent and multi-agent frameworks. In: *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence.* 2019; 286–292.
36. Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerström test for nicotine dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br J Addict.* 1991;86(9):1119–1127.
37. Mushtaq N, Beebe LA. Assessment of the tobacco dependence screener among smokeless tobacco users. *Nicotine Tob Res.* 2016;18(5):885–891.
38. Oh YJ, Zhang J, Fang M-L, Fukuoka Y. A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss. *Int J Behav Nutr Phys Act.* 2021;18(1):1–25.
39. 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(10):e20346e20346.
40. Munafò MR, Nosek BA, Bishop DV, *et al.* A manifesto for reproducible science. *Nat Hum Behav.* 2017;1(1):1–9.
41. Gupta SK. Intention-to-treat concept: a review. *Perspect Clin Res.* 2011;2(3):109–112.
42. Reeves B, Nass C. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.* Cambridge, UK: Cambridge University Press. 1996;305.
43. Strecher V, McClure J, Alexander G, *et al.* The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. *J Med Internet Res.* 2008;10(5):e36e1002.
44. Mermelstein R, Cohen S, Lichtenstein E, Baer JS, Kamarck T. Social support and smoking cessation and maintenance. *J Consult Clin Psychol.* 1986;54(4):447–453.
45. Chaiton M, Diemert L, Cohen JE, *et al.* Estimating the number of quit attempts it takes to quit smoking successfully in a longitudinal cohort of smokers. *BMJ Open.* 2016;6(6):e011045.
46. Safi Z, Abd-Alrazaq A, Khalifa M, Househ M. Technical aspects of developing chatbots for medical applications: scoping review. *J Med Internet Res.* 2020;22(12):e19127.
47. Ma T, Sharifi H, Chattopadhyay D. Virtual humans in health-related interventions: a meta-analysis. Presented at CHI Conference on Human Factors in Computing Systems; May 4–9, 2019; Glasgow, Scotland, UK.
48. Bickmore T, Gruber A. Relational agents in clinical psychiatry. *Harv Rev Psychiatry.* 2010;18(2):119–130.
49. Miller WR, Moyers TB. The forest and the trees: relational and specific factors in addiction treatment. *Addiction.* 2015;110(3):401–413.
50. Bickmore TW, Mitchell SE, Jack BW, *et al.* Response to a relational agent by hospital patients with depressive symptoms. *Interact Comput.* 2010;22(4):289–298.

51. Joe GW, Simpson DD, Dansereau DF, Rowan-Szal GA. Relationships between counseling rapport and drug abuse treatment outcomes. *Psychiatr Serv*. 2001;52(9):1223–1229.
52. Bickmore T, Schulman D, Yin L. Maintaining engagement in long-term interventions with relational agents. *Appl Artif Intell*. 2010;24(6):648–666.
53. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. *Transl Behav Med*. 2017;7(2):254–267.
54. Abd-Alrazaq AA, Alajlani M, Alalwan AA, et al. An overview of the features of chatbots in mental health: a scoping review. *Int J Med Inform*. 2019;132:103978.
55. Fadhil A. Can a chatbot determine my diet?: Addressing challenges of chatbot application for meal recommendation. *arXiv*. Preprint posted online on February 25, 2018.
56. Abd-Alrazaq AA, Alajlani M, Ali N, et al. Perceptions and opinions of patients about mental health chatbots: scoping review. *J Med Internet Res*. 2021;23(1):e17828.
57. Car LT, Dhinakaran DA, Kyaw BM, et al. Conversational agents in health care: scoping review and conceptual analysis. *J Med Internet Res*. 2020;22(8):e17158.
58. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. *J Med Internet Res*. 2016;18(10):e264e6349.