Effectiveness and Acceptability of Conversational Agents for Sexual Health Promotion: A Systematic Review and Meta-Analysis

Divyaa Balaji¹, Linwei He², Stefano Giani³, Tibor Bosse⁴, Reinout Wiers⁵, Gert-Jan de Bruijn⁶

¹Amsterdam School for Communication Research, University of Amsterdam, Amsterdam, Netherlands
²Department of Communication and Cognition, Tilburg University, Tilburg, Netherlands
³University Library, University of Amsterdam, Amsterdam, Netherlands
⁴Behavioural Science Institute, Radboud University, Nijmegen, Netherlands
⁵Department of Psychology, University of Amsterdam, Amsterdam, Netherlands
⁶Department of Communication Science, University of Antwerp, Antwerp, Belgium
Abstract

Digital health interventions for sexual health promotion have evolved considerably alongside innovations in technology. Despite these efforts, studies have shown that they do not consistently result in the desired sexual health outcomes. This could be attributed to low levels of user engagement, which can hinder intervention effectiveness as users do not engage with the system enough to be exposed to the intervention components. It has been suggested that conversational agents have the potential to overcome the limitations of prior systems and promote user engagement through the increased interactivity offered by bidirectional, natural language-based interactions. The present review therefore provides an overview of the effectiveness and user acceptability of conversational agents for sexual health promotion. A systematic search of seven databases provided 4,534 records and after screening, 31 articles were included in this review. A narrative synthesis of results was conducted for effectiveness and acceptability outcomes, with the former supplemented by a meta-analysis conducted on a subset of studies. Findings provide preliminary support for the effectiveness of conversational agents for promoting sexual health, particularly treatment adherence. These conversational agents were found to be easy to use and useful, and importantly resulted in high levels of satisfaction, use and intentions to reuse, while user evaluations regarding the quality of information left room for improvement. The results can inform subsequent efforts to design and evaluate these interventions, and offer insight into additional user experience constructs identified outside of current technology acceptance models which can be incorporated into future theoretical developments.

Keywords: sexual health, review, meta-analysis, HIV, chatbot, conversational agent, mHealth, digital health intervention
Introduction

Digital health interventions (DHIs) are interventions delivered via digital technologies such as smartphones, websites, social media or text-messaging. DHIs have become increasingly popular for large-scale health promotion efforts as an innovative, cost-effective, and scalable solution for addressing key public health challenges such as staff shortages and budget constraints. They can be especially suited for the sexual health domain because there is the potential to reach at-risk groups e.g. adolescents, ethnic minorities, sexual minorities, illicit drug-users and sex workers are less likely to seek professional care due to limited resources, poor quality of services and stigmatization.

DHIs for sexual health promotion have evolved considerably over time alongside increased internet and mobile device adoption as well as new technologies. Early DHIs leveraging short-message service (SMS) and digital media e.g. websites, video and CD-ROM became widespread and were well-received by users. However, positive results were observed largely for sexual health knowledge and attitudes and less so for behaviours such as treatment adherence, human papilloma virus (HPV) vaccine uptake and condom use. Moreover, intervention effectiveness tended to decline over time. Smartphones, in spite of their pervasiveness and functionalities, failed to gain positive responses from the community with regards to mobile phone applications intended to promote STI prevention and care. The use of emerging technologies such as serious gaming, virtual reality and social media faced resulted in similar outcomes, whereby little to no effects were observed for most sexual health behaviours with the exception of testing uptake. Ironically, while the motivation behind the dynamic adoption of technological innovations was likely increased user engagement, several authors have attributed the observed low intervention effectiveness to reduced immersion and flow i.e., feeling fully involved and focused on the activity. This is
consistent with the notion that positive user engagement precedes positive interactions with intervention components, thereby leading to increased intervention efficacy. Continuing in the footsteps of adopting new technologies, there has been a recent shift towards conversational agents (CAs) for delivering DHIs across healthcare domains such as substance abuse, mental health, exercise and even stress-reduction. For the purposes of this review, we adopt the definition of CAs as systems that can simulate conversation with users through natural language such as written text or voice thus permitting automated two-way communication between the user and system. Examples of CAs range from the well-known open-domain virtual voice assistants such as Siri and Alexa to customer service chatbots available through commercial websites and social media platforms such as Facebook and even embodied CAs which employ computer-generated avatars. It has been suggested that DHIs allowing two-way interactions can increase intervention efficacy by addressing both non-intentional and intentional forms of non-adherence to target health behaviours, and promoting user engagement by encouraging users to explore their attitudes and feelings in a more productive and personally relevant manner. Furthermore, CAs particularly hold promise as a more innovative way to communicate with younger users given their high digital literacy and familiarity with chat applications. Given the rate at which the field of natural language processing is advancing, CAs can also increase engagement by understanding the user and providing intelligent, relevant communication at all times to different target populations.

In essence, it appears relevant to already examine the prospects that CA-based DHIs may offer for sexual health promotion. Furthermore, prior reviews on DHIs in this domain have primarily addressed effectiveness, often with emphasis on randomized controlled trials or peer-reviewed literature. However, this approach may result in the exclusion of studies which could provide insight into the potential of CA-based interventions by...
assessing user acceptability or reporting preliminary findings given the rapid advances in conversational technology. As the effectiveness and user acceptability of DHIs are inextricably linked, the aim of the present systematic review is to summarize available evidence regarding both the effectiveness and acceptability of conversational agents for sexual health promotion.
Method

Literature Search

A systematic search of the literature was performed in October 2020 using PsycINFO, Web of Science, ACM Digital Library, IEEE Xplore, Medline, Embase and CMMC, not restricted by publication year or language. An updated search was performed in February 2021. Grey literature identified in those databases, including dissertations, theses, and conference proceedings, were also included for screening given the infancy of this field. Two sets of search terms were devised by the first reviewer (DB) and a librarian (SG) and customized for each selected database. The first set addressed conversational agents and included other related terms such as ‘chatbots’, ‘relational agent’, ‘virtual assistant’, ‘dialog system’ and ‘mHealth’. The second set addressed sexual health and included other related terms such as ‘HIV’, ‘sexually transmitted diseases’, ‘HPV’ and ‘syphilis’. The full search strategy can be found in the Supplementary Material. The protocol for this review was registered at the International Prospective Register of Systematic Review (PROSPERO; registration number CRD42021222969).

Inclusion and Exclusion Criteria

Included studies had to meet three criteria: (i) described conversational agents or synonymous systems that permitted two-way interactions that were fully automated (i.e. without any human mediation)\(^1\), (ii) addressed any sexually-transmitted condition such as HIV/AIDS, HPV and other STIs or targeted aspects of sexual health promotion such as medication adherence and reducing risky sexual behaviour and (iii) described an evaluation applied to the technology, focusing on either health outcomes or end-user evaluations (e.g. acceptability, usability, or satisfaction, but not cost-effectiveness and cost-analysis outcomes). Forms of one-way

\(^{1}\) Also includes systems that may not intuitively be considered CAs, namely automated two-way text-messaging systems – they can be considered “old school” CAs that use SMS instead of chat applications.
communication, human-mediated communication and systems without response contingency (e.g. quizzes, ecological momentary assessment and computer-assisted self-interviewing) were excluded, as were studies which only evaluated the idea or content that will later be implemented, as the actual technology does not undergo evaluation. Reviews, meta-analyses, protocol papers and poster abstracts were excluded, as were citations with missing abstracts. No restriction was placed on study design given the varied and dynamic nature of the field.

Screening procedures were piloted by the first and second reviewers (DB and LH). Once a sufficient kappa value was achieved (>0.6, indicating substantial inter-rater agreement), the initial screening of articles was conducted independently based on the information contained in the title and abstract. Conflicts were discussed between DB and LH, and unresolved conflicts were discussed with a third reviewer (GJ). The same procedure was applied for full-text screening. After this, citation tracking was conducted to ensure that all relevant studies were identified, resulting in 31 included studies (Figure 1).

Data Extraction and Synthesis

Data extraction was conducted by DB and reviewed by GJ for completeness. The following data from the included studies were extracted: title, author, year, study design, sample size, target health behaviour, target population, interaction frequency, intervention duration, theoretical framework, technology platform and initiator and outcomes. Outcomes were classified as either health outcomes or user evaluations and were all summarized through a vote-counting strategy and are presented as a narrative synthesis of results. Health outcomes assessed in more than one study using a randomized controlled trial design were additionally analyzed through a meta-analysis. User evaluation outcomes, quantitative and qualitative, were organized around the components of a technology acceptance model (TAM), namely the DeLone and McLean Information Systems Success Model (D&M Model) - system quality, information quality, user satisfaction, actual use & intent to reuse and net benefits.
136 **Meta-analysis**

Health outcomes assessed in more than one study using a randomized controlled trial design were additionally analyzed through a meta-analysis. For the meta-analysis, random effect models were used. Per included study, we calculated the effect size $d$ and its standard error using suggested formulas for mean differences and odds ratio \(^{52}\) in MacOS Numbers. For studies that had multiple outcome measures, an average effect size $d$ and standard error was calculated using suggested formulas\(^{52}\) and setting the multiple outcome correlation as $r = .50$.

Heterogeneity was derived from the Q-statistic\(^{53}\) and publication bias from funnel plots and Egger’s test\(^{54}\). When publication bias was present, trim-and-fill analysis was conducted \(^{55}\). We conducted the meta-analysis for the five RCT studies including all outcomes, all objective outcomes, and all objective ART outcomes respectively (so excluding\(^{56}\)) with the dmetar, meta, and metafor packages in RStudio (version 2021.09.1) for MacOS.
Results

Overview of Studies

Table 1 provides a summary of the included studies (n = 31). Thirteen RCTs were identified, and the remaining utilized either pre-post or post-study designs. Social Cognitive Theory was cited most often as the underlying theoretical framework, followed by the Information-Motivation-Behavioral Skills model and Motivational Interviewing. Seventeen out of 31 studies did not indicate an underlying theoretical framework.

The most common target health behaviour was antiretroviral therapy (ART) adherence (15/31). In contrast to ART, only four studies focused on preventive treatment i.e. pre-exposure prophylaxis (PrEP) and post-exposure prophylaxis (PEP) knowledge, uptake and/or treatment adherence. Other studies addressed general sexual health and safe sex practices, HIV/STI testing and HIV risk factors such as heavy drinking and medication adherence for bipolar disorder. HPV vaccine attitude and uptake was the subject of three studies, targeting either mothers or young adults. Studies predominantly took place in the United States. The populations of interest were broadly categorized as individuals at risk for HIV and HIV-positive individuals who are initiating or taking ART, particularly those who have sub-par adherence. Within these, groups which were targeted often were young people, ethnic minorities, sexual minorities, women and substance users.

Mobile phone was the most common platform to deliver the interventions, which included short-message service (SMS), interactive voice response (IVR), instant messaging (IM) and smartphone application (app). The remaining interventions were delivered via computer, either through websites or desktop/tablet applications. While all of the included systems permitted two-way interactions,
seventeen out of 31 studies allowed multi-turn interactions i.e., either the CA or user was able to respond more than once within the same conversation\(^{57,58,62,65–70,72,74,75,77–81}\) and the remainder which were one-turn-only interactions \(^{56,59–61,63,64,71,73,76,82–86}\). Systems generally initiated interactions with users at fixed times e.g. daily \(^{58,60–64,73,76,77,82,84–86}\), twice a week \(^{57}\), thrice a week \(^{67,68}\), weekly \(^{56,60,71,83}\) and monthly \(^{70}\). Some were user-initiated \(^{59,65,69,74}\) while the rest were only used once to assess user acceptability outcomes.

Out of 31 studies, 26 looked at one or more user acceptability outcomes and 20 studies addressed intervention effectiveness through appropriate health outcomes.

**Effectiveness**

Out of 11 studies that assessed antiretroviral therapy outcomes, studies which used self-report measures \(^{62,65,70,77,82,84,86}\) found support for intervention effectiveness whereas those employing pill count \(^{63–65,77,82,86}\) and biomarker measures \(^{62,67,70,82–84}\) did not support this. Pre-exposure prophylaxis outcomes were assessed in three studies also using self-report \(^{71}\), pill count \(^{71}\) and biomarker measures \(^{56,73}\), all of which indicated significant effects. Positive findings were observed for multi-dimensional attitude towards HPV vaccine in two studies \(^{59,78}\), of which one found an additional near-significant effect for HPV vaccine uptake behaviour \(^{59}\). For condom use, two studies found no significant effects for condom use behaviour \(^{57,68}\), of which one found an improvement in condom use attitude but not intention \(^{68}\). One intervention was targeted at promoting a range of safe sex practices for women during pre-conception and was found to be effective for reducing the number of risks associated with (sexually transmitted) infectious diseases \(^{69}\).
Meta-analysis

The included RCTs looked at medication adherence for either antiretroviral therapy \textsuperscript{62,67,70,77} or pre-exposure prophylaxis\textsuperscript{56}. Table 2 summarizes the results for the conducted meta-analyses. For the five RCTs \((n = 582)\), there was a small and significant effect on medication adherence, \(d^+ = 0.23, 95\% \text{ CI} [0.037; 0.4213], p = .030\) favoring the chatbot intervention (Figure 2). This effect was not present after repeating the analysis for only objective outcomes \((d^+ = 0.19, 95\% \text{ CI} [-0.096; 0.484], p = .137)\) (Figure 3), and with only ART studies \((d^+ = 0.10, 95\% \text{ CI} [-0.232; 0.424], p = .419)\) (Figure 4). The funnel plots did not indicate publication bias (see Figures 5, 6 and 7).

User Acceptability

System Quality

Nine out of 26 studies evaluated system quality\textsuperscript{60,65,69,72,75,79,85,86}. While ease of use was the most common measure (7/10), system quality was also evaluated through overall usability\textsuperscript{79}, pragmatic quality\textsuperscript{72} and response speed\textsuperscript{75,78}. Users generally found the systems straightforward and easy to use. This was attributed to technological capabilities such as quick replies in instant messaging\textsuperscript{75} and the general familiarity of mobile phone interfaces which ensured that the chatbot understood them. Interestingly, response speed was evaluated negatively in two studies, whereby users found it to be either unrealistically fast\textsuperscript{75} or too slow and not efficient enough for a machine\textsuperscript{78}.

Information Quality

Information quality was assessed in twelve out of 26 studies along content relevance\textsuperscript{58,66–69}, content quantity\textsuperscript{58,66,85}, clarity\textsuperscript{66,78,81}, language style\textsuperscript{58,66,75}, interaction intensity (depth, frequency and duration)\textsuperscript{58,61,62,68} and repetitiveness\textsuperscript{67,78,85}. While participants largely found the content relevant and useful for the target behaviour\textsuperscript{57,58,68,69} (e.g. “these questions
are things that all girls think about…made me think about my behaviours\(^\text{57,}\)\), they voiced the desire for additional personalization of content\(^\text{67,68}\). Across three studies that assessed content quantity \(^\text{58,66,85}\), two reported that users wanted additional content on other health topics (e.g. side effects of ART\(^\text{66}\)) and on non-health topics (e.g. communication and relationships \(^\text{58}\)).

Language style was evaluated in three studies, two of which found that language style should be more appropriate. For example, it is important to avoid sensitive phrases (e.g. using AIDS interchangeably with HIV) as well as graphic images portraying sickness\(^\text{66}\). For voice-based systems, users discouraged any harsh or judgmental intonation and wanted “straight talk” like from friend or relative\(^\text{58,66}\). Van Heerden and colleagues \(^\text{75}\) instead found that users thought the language was too formal and incongruent with real life conversations. Qualitative feedback indicated that the clarity of some systems can be further improved by rewording and using visuals to complement the verbal and/or audio-visual dialogue\(^\text{66,78,81}\). Studies looking at repetitiveness \(^\text{67,78,85}\) found that some users were irritated when the system did not exhibit the variety that is characteristic of natural conversations. Users expressed room for improvement regarding interaction intensity in four studies \(^\text{58,61,62,68}\) – they wanted the conversations to occur more often i.e. daily or more and last longer \(^\text{58,61}\) while others would have appreciated more or less messages sent to them depending on their preferences \(^\text{62,68}\).

**User Satisfaction**

Nine out of 26 studies assessed overall user satisfaction. Two studies \(^\text{68,70}\) made use of the Client Satisfaction Questionnaire while others enquired about satisfaction \(^\text{59,62,64,85,86}\) or acceptability \(^\text{57,81}\) using one or more survey items. Across all studies, users reported above-average scores and that they found their experiences with the technology satisfying or enjoyable.
Use

*Intent to reuse* the system in the future was assessed in eight studies and most users (around 78% across studies) responded positively, expressing that they would like to continue receiving the intervention after the study or be open to using such a system in their daily lives.

*Actual use* was assessed quantitatively in 10 studies through message response rate for SMS-based systems and through usage metrics for smartphone applications and websites. The average message response rate was around 69%, ranging from 47% to 68% on the lower end to 92% on the higher end. Usage metrics indicated that an average of 88% participants accessed the systems at least once and each interaction lasted around 10 minutes, and another system received 4,390 topic-relevant messages with an average of three questions per session, indicating reasonable use of the system.

Net Benefits

Perceived net benefits were evaluated in more than half of the studies (14/26) through perceived usefulness and the likelihood of recommending the system to other individuals. Qualitative feedback from users indicated that the systems were useful for promoting a range of sexual health behaviours such as condom use (“these questions are things that all girls think about…”), HIV testing (“it could save time not having to wait at a clinic for a counsellor”) and HPV vaccine uptake (“…provided useful information and reinforced important points”). Some studies found that the systems targeting treatment adherence were only useful if users were facing difficulties with adequate adherence, consistent with the otherwise positive evaluations of perceived usefulness regarding systems targeting either individuals initiating treatment or exhibiting poor adherence. Reminders were cited as being most useful feature by providing different strategies, minimizing forgetfulness when they were busy or at work and that the reminders did not stop until they texted back. Overall, users were also highly likely (86% across all
studies) to recommend the systems they used to others who are HIV-positive, to a friend, or to others in general.

Additional User Acceptability Outcomes

There were constructs identified in the set of user acceptability outcomes that did not fall under any component of the D&M Model and are therefore summarized below. Constructs associated with privacy and anonymity were assessed in five studies, whereby a minority of users voiced a desire for additional measures (e.g. the ability to hide the application screen quickly, minimize attention from alerts and reminders) to avoid unintentional disclosure in three studies. Two studies assessed trust in the system and received positive feedback from users. Questions regarding general feelings of comfort and the lack of stigma were administered to users in five studies out of which four revealed that users indeed felt safe and comfortable in their interactions with the system. One study found that a small number of users faced instances of potential embarrassment and stigma when using the system in public and near their friends. In five studies, users were asked about the extent to which they felt emotionally supported, or cared for, by the system, all of which reported positive findings. Out of the seven studies that looked at social presence, users expressed desire for increased social presence or actual human interaction in three studies while the remainder, most of which utilized a static or embodied avatar, reported good or sufficient social presence.
Digital health interventions (DHIs) for sexual health promotion are becoming increasingly commonplace and are particularly attractive because at-risk groups are often unable to or are reluctant to seek out professional advice. In the spirit of adopting new technologies, DHIs using conversational agents (CAs) have begun to receive more attention for their added capacity to imitate natural interactions with humans. The CAs included in this review exhibited a marked variety in how the technology interacted with the users. Of interest were the relatively large number of CAs that allowed multi-turn interactions, which come across as more natural and are characterized by increased interactivity and feedback. These were particularly pronounced in the more recent years, likely explained by the rapid technological advances that have been made in the field of artificial intelligence. Understandably, the included studies were largely pilot studies which indicated the infancy of this growing field but resulted in the lack of rigorous study designs utilizing appropriate control groups that would have aided in more empirical analysis. As the interest in CAs for sexual health is evidently rising, the present review situates itself well in summarizing the available evidence of their effectiveness and acceptability.

Most of the studies targeted medication adherence, either antiretroviral therapy (ART) or pre-exposure prophylaxis (PrEP), and meta-analytic findings supported the effectiveness of CA interventions for adherence when considering both self-report and objective measures as well as both medications. However, this result became non-significant upon excluding self-report measures and the single PrEP study, suggesting the need to consider how these systems can result in more tangible improvements for ART adherence. In general, CAs targeting ART adherence fall under the umbrella of treatment and were able to help individuals who already have HIV to manage their symptoms through interactive reminders and information provision. In contrast, there were markedly fewer interventions addressing prevention of sexual health-
related diseases, namely pre-exposure prophylaxis (PrEP), condom use, human papilloma virus (HPV) vaccine uptake and sexually-transmitted disease (STD) testing. Based on this limited number of studies, CA interventions resulted in positive outcomes for attitudes towards condom use and HPV vaccination uptake and testing behaviour. Given the difficulties in getting individuals to engage in precautionary behaviours, additional studies are needed to explore how CAs can effectively support such behaviours.

Through the lens of the DeLone and McLean Information System Success Model, the CA interventions were found to be acceptable to users in terms of ease of use, perceived usefulness, satisfaction and intent to reuse, which is likely to translate into actual use according to the D&M model and other technology acceptance models. In addition to reminders, CAs were seen as being capable of providing on-demand emotional support and useful information in an anonymous manner without human contact, consistent with proposed drivers of CAs for healthcare. However, they were found to be lacking in aspects of information quality such that users desired additional, personalized, and clearer content communicated in an appropriate language style. While not included in traditional technology acceptance models as a separate construct, the heavy reliance on textual content and communication principles in conversational systems lends to the importance of information quality, a point that is further supported by the number of included studies that evaluated this construct in some form.

The study identified an additional set of constructs that may play a role in user acceptability within this domain. Trust and privacy can be thought of as contemporary challenges that have permeated emerging technologies and have been discussed in other extended TAM models within and outside of healthcare. Users both expected and were largely satisfied with the degree of privacy and trustworthiness exhibited by the systems, although the demand for security appears abundantly strong that additional features may be needed to
motivate long-term use. Social presence was found to be sufficient in only half of the studied systems, which could be attributed to the use of multimedia and embodiment through avatars\textsuperscript{43,107,108}. An interesting issue arises in incorporating social presence into a system that is often touted for its capacity to enable anonymous, “non-human” and therefore self-disclosing interactions\textsuperscript{45,109–111}, suggesting the need to achieve a delicate balance. While social presence has been implicated in the user acceptance of conversational agents in other areas\textsuperscript{38,108,112,113,114,115,116,117}, its role in the sexual health domain remains to be disentangled. Comfort and emotional support have not received as much attention in the literature but these findings suggest that they may be important in specific domains such as domestic violence\textsuperscript{118,119}, sexual issues\textsuperscript{120,121} and mental health\textsuperscript{122} where individuals need to feel safe and accepted while engaging with the system. While this review underscores the potential importance of these constructs, future studies can explore their role in user acceptance and inform their inclusion in extended TAM models for sexual health and related domains.

**Conclusion**

Despite the limited body of evidence, these findings support the notion that CAs for sexual health may not only be effective but that users also find these useful and acceptable for a range of sexual health behaviours. While CAs are already capable of supporting antiretroviral therapy adherence through simple two-way interactions, more studies are required to understand how the potential of CAs can be leveraged for more complex behaviours. This review also emphasizes the value of rigorous, holistic, and mixed-method evaluations of CA-based DHIs to gain deeper insight into how the intervention components are perceived by users as a driver of intervention efficacy. To that end, the findings serve as a good starting point for how we might go about enhancing the user experience for these interventions and highlight the need for theoretical developments regarding technology acceptance models which are more applicable to sensitive domains. The question still remains for further research
as to whether and under what circumstances individuals would voluntarily adopt CAs outside the research context and in what way they can be reached in practice.
Declarations

Funding

This work was supported by The Netherlands Organization for Scientific Research (NWO) [grant number 406.D1.19.054].

Data Availability

The authors confirm that the search strategy used in this study is available within the article and/or its supplementary materials. The data supporting this systematic review and meta-analysis are from previously reported studies and datasets, which have been cited. The processed data are available from the corresponding author upon request.

Conflict of Interest

The authors declare no conflicts of interest.
References


58. Davis T, Ralph ;, Diemonte J, Prietula M. Using ADAPT-ITT to Modify a Telephone-Based HIV Prevention Intervention for SMS Delivery: Formative Study. doi:10.2196/22485

59. Pot M, Paulussen TG, Ruiter RA, et al. Effectiveness of a Web-Based Tailored Intervention With Virtual Assistants Promoting the Acceptability of HPV Vaccination Among Mothers of


84. Dowshen N, Kuhns LM, Johnson A, Holoyda BJ, Garofalo R. Improving Adherence to Antiretroviral Therapy for Youth Living with HIV/AIDS: A Pilot Study Using Personalized,


