

Conversational Agents for Chronic Disease Self-Management: A Systematic Review

Ashley C. Griffin, MSPH¹⁻², Zhaopeng Xing, MPS¹⁻², Saif Khairat, PhD, MPH, FAMIA¹⁻²,
Yue Wang, PhD¹⁻², Stacy Bailey, PhD, MPH³, Jaime Arguello, PhD, MS¹,
Arlene E. Chung, MD, MHA, MMCI, FAMIA^{1-2,4}

¹University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; ²Carolina Health Informatics Program, Chapel Hill, NC, USA; ³Northwestern University, Chicago, IL, USA; ⁴Program on Health & Clinical Informatics, Chapel Hill, NC, USA

Abstract

We conducted a systematic literature review to assess how conversational agents have been used to facilitate chronic disease self-management. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework was used. Literature was searched across five databases, and we included full-text articles that contained primary research findings for text-based conversational agents focused on self-management for chronic diseases in adults. 1,606 studies were identified, and 12 met inclusion criteria. Outcomes were largely focused on usability of conversational agents, and participants mostly reported positive attitudes with some concerns related to privacy and shallow content. In several studies, there were improvements on the Patient Health Questionnaire ($p < 0.05$), Generalized Anxiety Disorder Scale ($p = 0.004$), Perceived Stress Scale ($p = 0.048$), Flourishing Scale ($p = 0.032$), and Overall Anxiety Severity and Impairment Scale ($p < 0.05$). There is early evidence that suggests conversational agents are acceptable, usable, and may be effective in supporting self-management, particularly for mental health.

Introduction

Sixty percent of U.S. adults have chronic diseases, such as hypertension and diabetes¹, which are 86%² of the \$2.6 trillion in annual health care expenditures in the U.S.³ Chronic diseases can be controlled through self-management, which refers to an individual's ability to manage symptoms, treatments, physical and psychosocial consequences of diseases, and lifestyle changes⁴. Successful self-management requires sufficient knowledge of the disease and the necessary skills to manage and prevent complications, slow disease progression, and improve health outcomes⁴. Given the near ubiquity of mobile phones with 96% of U.S. adults owning a mobile phone and 81% owning a smartphone in 2019⁵, mobile health (mHealth) approaches to facilitate self-management are a promising supplement or alternative to traditional programs. In particular, short message service (SMS) interventions have shown improvements across a range of self-management behaviors, such as medication adherence, smoking cessation, and physical activity⁶⁻⁹. Patients have reported SMS apps to be highly acceptable and easy to use, though receiving tailored text messages in real-time could improve the user experience^{10,11}. This suggests that conversational interfaces, which can provide automated two-way communication and evidence-based tailored responses, may have potential advantages over current self-management interventions.

Conversational agents, also known as dialogue systems or chatbots, are systems that communicate with users in natural language (text or speech) and can complete specific tasks or mimic "chat" characteristics of human-human interactions¹². Conversational agents have the ability to mirror a therapeutic process, such as cognitive behavioral therapy or motivational interviewing¹³. This could promote goal setting, positive feedback, self-monitoring, overcoming obstacles to self-management, and education, which are important components in chronic disease self-management. Text-based conversational agents can engage users about sensitive or stigmatized topics and have the ability to show empathy, which can help to alleviate negative symptoms or emotions^{14,15}. For example, a conversational agent could be paired with wearable devices to detect the onset of a disease exacerbation and provide the user with appropriate therapy to help manage their symptoms¹⁶.

Conversational syntax and semantics could also reveal rich linguistic information about the user, which may allow for a better understanding of the person's health status¹⁵. Given the potential to deliver self-management components through a conversational agent, the objective of this systematic review was to assess how text-based conversational agents have been used for chronic disease self-management.

Methods

Search Strategy. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were used as a framework¹⁷. The literature search was conducted within five electronic databases (PubMed, Scopus, Embase, Cumulative Index to Nursing and Allied Health Literature, and Association for Computing Machinery Digital Library). We prioritized sensitivity over specificity to ensure that all potentially relevant studies were included. The search was limited to papers published prior to the search date of October 1, 2018. Search terms included variations of: “conversational agent, virtual coach, chatbot, dialogue system, and health.” Search terms could appear anywhere in the title, abstract, or keywords. Bibliographies of included publications were also searched to identify additional relevant literature. All studies were imported from EndNote X8 into Covidence systematic review software to screen studies. Duplicate studies across databases were removed.

Eligibility Criteria. The seven inclusion criteria were: (1) full-text journal articles or conference proceedings, (2) published in English, (3) contained primary research findings, and (4) included a text-based conversational agent in (5) the context of self-management for (6) chronic diseases in (7) adults 18 years and older. Editorials, letters, design papers, conference abstracts, and study protocols were excluded. For this study, “conversational agent” was defined as an autonomous system that can communicate with users bi-directionally in text¹². Thus, automated text messages for reminders or appointments and question and answer systems were not considered conversational agents. Since the focus was on text-based conversational agents, only studies that contained agents with both text-based input and output were included, though the system could also contain speech-based input or output. “Wizard of Oz” studies, which involved a human simulating the response of the system, were included since these studies could potentially reveal valuable information about the nature of interaction and functionalities of the system in relation to self-management.

Self-management was defined as an individual’s ability to manage symptoms, treatments, physical and psychosocial consequences, and lifestyle changes⁴. We adapted the list of self-management skills from Lorig et al (2003), which has been used in numerous studies and in the Stanford Chronic Disease Self-Management Program^{18,19}. Studies that did not contain at least one of these self-management skills were excluded. We included conversational agents targeted towards one or more of the 19 chronic diseases defined by the Centers for Medicare and Medicaid Services²⁰. Because chronic diseases are defined as conditions that last longer than one year, we also included additional prevalent conditions and populations where self-management is important that were not in the Centers for Medicare and Medicaid Services list, such as obesity and mental health conditions (post-traumatic stress disorder, substance use/addiction). We also included studies for older adults who often have chronic diseases.

Study Selection. Two reviewers (AG, ZX) independently rated each title and abstract as “potentially relevant” or “not relevant” based upon the abstract potentially meeting the inclusion criteria. Then, each reviewer examined the full-text of the studies rated as “potentially relevant” and applied the inclusion criteria again. Any disagreements were discussed with an adjudicator (AC). Cohen’s kappa statistic²¹ was calculated to measure inter-rater agreement between each reviewer for screening and full-text review.

Data Extraction and Synthesis. One reviewer (AG) extracted data from the abstract, main text, and supplemental material into a standardized form, and another reviewer (ZX) examined the form for accuracy. Chronic diseases were grouped into mutually exclusive categories based on the primary disease or population the conversational agent was targeted towards. Self-management skills and attributes of the conversational agent were grouped into one or more categories based on the description of the conversational agent provided in the study. The extracted data were cross tabulated to show a comparison of key concepts across studies.

Risk of Bias. We assessed the risk of bias within each study using the Agency for Healthcare Research and Quality internal validity quality evaluation form²². One reviewer (AG) evaluated and rated each study’s risk of bias as good, fair, or poor, and another reviewer (ZX) reviewed the form for consistency. We did not assess the validity of the literacy measurement used in each study as there was limited data across studies²².

Results

Study Selection. 2,422 studies were identified from the database search, and 1,606 studies were screened after removing duplicates. 82 studies were included in the full-text review, and 12 studies were included in the final review, which was comprised of 12 unique conversational agents (see Figure 1). The kappa statistic for title and abstract screening was 0.94 (strong agreement) and 0.75 (moderate agreement) for full-text screening.

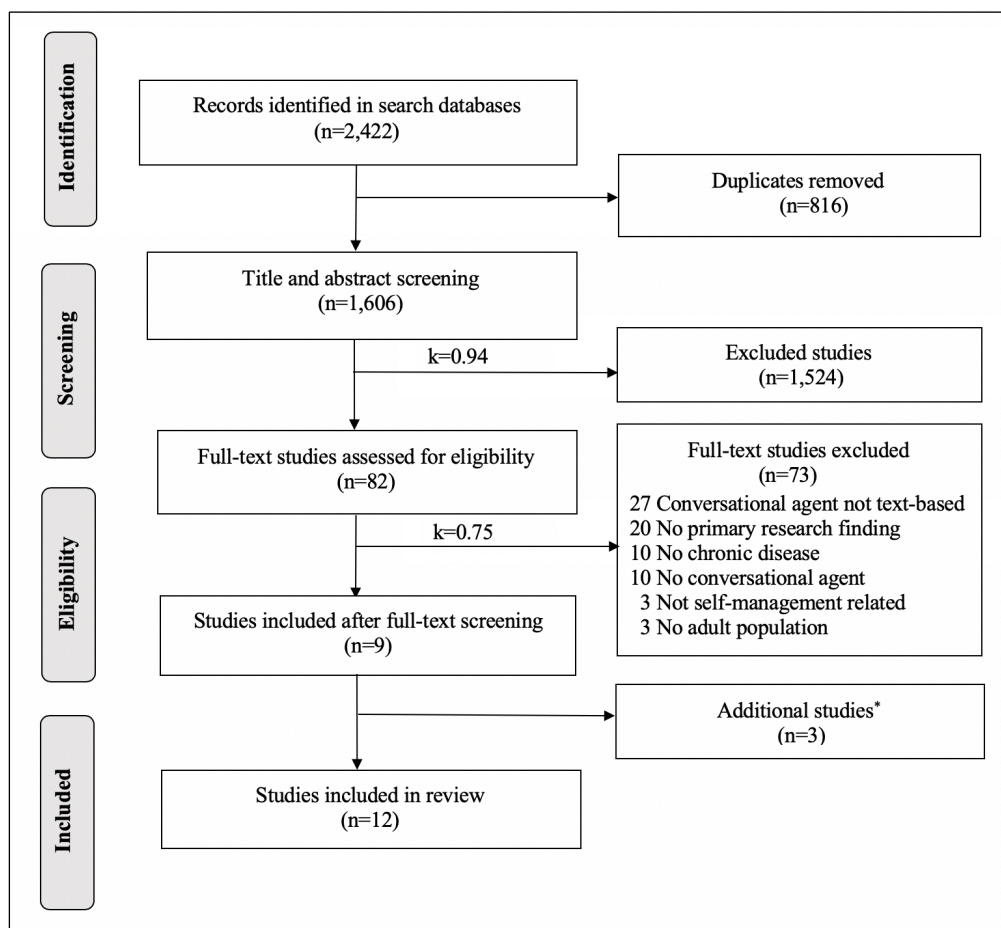


Figure 1. PRISMA diagram.

*The three studies were included through a review of reference citations from primary studies.

Study Characteristics. The final review included studies²³⁻³⁴ published from 2012-2018 from seven countries. There was a range of study designs and methodologies, including five randomized controlled trials, five quasi-experimental studies, and two non-experimental studies (see Table 1). The majority were pilot studies and study duration ranged from two weeks to four months. The total number of participants ranged from 10 to 401, and participant ages were between 18 and 92 years.

Table 1. Study methodology and outcomes of reviewed studies.

Study	Study Design and Methodology	Participant Characteristics	Outcomes
Baskar et al, 2015 ²³	Quasi-experimental + interviews 11 older adults and 2 health professionals played the role of a fictitious persona while interacting with the agent for ~10 minutes followed by an interview.	Age: N/A Gender: 55% female Race: N/A Baseline Clinical Characteristics: N/A	<ul style="list-style-type: none"> • Attitudes, perceptions • Usage
Elmasri et al, 2016 ²⁴	Quasi-experimental + interviews 17 participants interacted with the agent to explore their alcohol consumption for ~10 minutes followed by a survey and interview.	Age: 18-25 Gender: 41% female Race: N/A Baseline Clinical Characteristics: <5 drinks/day	<ul style="list-style-type: none"> • Attitudes, perceptions • Client Satisfaction Survey • Usage

Table 1. Study methodology and outcomes of reviewed studies, continued.

Study	Study Design and Methodology	Participant Characteristics	Outcomes
Fitzpatrick et al, 2017 ²⁶	<p>RCT</p> <p>70 participants were randomly assigned to Woebot or directed to a National Institute of Mental Health ebook for 2 weeks. Surveys were completed at baseline and post-intervention.</p>	<p>Age: 18-28</p> <p>Gender: 67% female</p> <p>Race: 79% Caucasian</p> <p>Baseline Clinical Characteristics: 46% had moderately-severe or severe PHQ-9 depression scores; 74% had severe GAD-7 anxiety scores</p>	<ul style="list-style-type: none"> • Patient Health Questionnaire (PHQ)-9 (p=0.017) • General Anxiety Disorder (GAD)-7 Scale (p=0.004) • Positive Affect Schedule (p=0.951) • Negative Affect Schedule (p=0.80) • Satisfaction (p<0.001) • Usage • Knowledge
Gaffney et al, 2013 ²⁵	<p>RCT</p> <p>48 participants were randomly assigned to MYLO or ELIZA to discuss a current problem for ~20 minutes. Surveys were completed at baseline, post-intervention, and 2-week follow-up.</p>	<p>Age: 18-32</p> <p>Gender: 79% female</p> <p>Race: N/A</p> <p>Baseline Clinical Characteristics: N/A</p>	<ul style="list-style-type: none"> • Resolution ratings (p<0.05) • Helpfulness ratings (p<0.05) • Depression and Anxiety Stress Scale (p=0.36) • Distress (p=0.13) • Usage
Kazemi et al, 2014 ²⁷	<p>Focus groups</p> <p>26 participants were placed into one of four focus group sessions to determine their views of mHealth technology to deliver an alcohol-related intervention.</p>	<p>Age: 18-20</p> <p>Gender: 73% female</p> <p>Race: 70% Caucasian</p> <p>Baseline Clinical Characteristics: N/A</p>	<ul style="list-style-type: none"> • Attitudes, perceptions
Ly et al, 2017 ²⁸	<p>RCT + interviews</p> <p>28 participants were randomly assigned to Shim or a wait list control group for 2 weeks. Surveys were completed at baseline and post-intervention. 9 participants from the intervention group were selected for a semi-structured interview.</p>	<p>Age: 20-49</p> <p>Gender: 54% female</p> <p>Race: N/A</p> <p>Baseline Clinical Characteristics: N/A</p>	<ul style="list-style-type: none"> • Flourishing Scale (p=0.032) • Perceived Stress Scale (p=0.048) • Satisfaction With Life Scale (p=0.10) • Attitudes, perceptions • Usage
Schroeder et al, 2018 ²⁹	<p>RCT</p> <p>84 participants were randomized into two messaging groups (semi-personalized messages or non-personalized messages). Participants completed weekly surveys over the 4-week study.</p>	<p>Age: 18-63</p> <p>Gender: 89% female</p> <p>Race: N/A</p> <p>Baseline Clinical Characteristics: 83% had an anxiety disorder on the Overall Anxiety Severity and Impairment Scale; 68% had moderate to severe range of depression on PHQ-9</p>	<ul style="list-style-type: none"> • Overall Anxiety Severity and Impairment Scale (p<0.05) • PHQ-9 (p<0.01) • System Usability Scale • Usage
Stein et al, 2017 ³⁰	<p>Quasi-experimental</p> <p>159 participants interacted with Lark for up to 16 weeks followed by a survey.</p>	<p>Age: 18-76</p> <p>Gender: 75% female</p> <p>Race: N/A</p> <p>Baseline Clinical Characteristics: BMI>25kg/m²</p>	<ul style="list-style-type: none"> • Weight loss • Healthy meals logged • Satisfaction • Usage

Table 1. Study methodology and outcomes of reviewed studies, continued.

Study	Study Design and Methodology	Participant Characteristics	Outcomes
Tsiourti et al, 2014 ³¹	Focus group + interviews 20 older adults and 14 health professionals participated in two focus groups and an interview to assess acceptance and expectations.	Age: 65-92 Gender: 65% female Race: N/A Baseline Clinical Characteristics: N/A	<ul style="list-style-type: none"> • Attitudes, perceptions
van Heerden et al, 2017 ³²	Quasi-experimental + interviews 10 Participants interacted with Lwazi/Nolwazi for ~25 minutes and provided feedback.	Age: 30 (average) Gender: 50% female Race: N/A Baseline Clinical Characteristics: N/A	<ul style="list-style-type: none"> • Attitudes, perceptions • Usage
Wang et al, 2018 ³³	Quasi-experimental 401 participants were placed into WeChat groups with an agent or received smoking cessation tips over the 8-week study. Participants completed weekly surveys.	Age: 33 (average) Gender: 40% female Race: N/A Baseline Clinical Characteristics: Smoked in the past 7 days	<ul style="list-style-type: none"> • Smoking cessation • Usage
Watson et al, 2012 ³⁴	RCT 70 participants were given a pedometer and randomly assigned to a conversational agent or access to website for 12 weeks. Surveys were completed at baseline and post-intervention.	Age: 42 (average) Gender: 84% female Race: 76% Caucasian Baseline Clinical Characteristics: BMI between 25-35kg/m ²	<ul style="list-style-type: none"> • Step count (p=0.07) • BMI (p=0.44) • Attitudes, perceptions • Usage

*Abbreviations: N/A: data not available in manuscript; RCT: randomized controlled trial; BMI: body mass index

Evaluation Measures. The majority of studies (10 of 12) evaluated usability, which was primarily measured by assessing users' attitudes towards using the conversational agent through questionnaires, interviews, or focus groups. Across the studies, the majority of participants reported moderate to high satisfaction with the agents, although only two studies used a validated questionnaire to assess satisfaction specifically. These two studies found the average satisfaction of the intervention was 3.6 out of 4 using the Client Satisfaction Survey²⁴ and 82 out of 100 on the System Usability Scale²⁹. Several studies incorporated the perceptions of healthcare professionals who expressed concerns with the large amount of information presented to older adults, cautioned use due to the potential to reduce independence, and stated the importance of the chatbot guiding rather than directing older adults^{23,31}. Ten studies evaluated usage of the agent, which was measured by the amount of time interacting with the agent or the number of sessions with the agent. For laboratory-based studies, participants interacted with the conversational agent on average between 10 to 26 minutes^{23-25,32}. For non-laboratory based experimental studies, the majority of participants interacted with the agent at least 50% of the study period^{26,28-30}. Seven studies evaluated clinical outcomes, which were primarily self-reported. There were significant improvements in the PHQ-9 (p<0.05)^{26,29}, GAD-7 (p=0.004)²⁶, Overall Anxiety Severity and Impairment Scale (p<0.05)²⁹, Flourishing Scale (p=0.032)²⁸, and Perceived Stress Scale (p=0.048)²⁸ among intervention groups. Only one study evaluated knowledge and found 100% of participants in the intervention group reported that they learned something new compared to 77% in the control group²⁶.

Conversational Agent Attributes. Conversational agents were focused on a variety of populations and chronic conditions such as depression^{25,26,28,29}, substance use^{24,27,33}, older adults^{23,31}, diabetes³⁰, overweight/obesity³⁴, and HIV/AIDS³² (Table 2). The majority of conversational agents were targeted towards mental health and assisted with managing symptoms or problem solving by helping users cope with their emotions and teaching them how to respond if their feelings worsened^{24-29,33}. For self-management skills, agents focused on skills such as maintaining a healthy

lifestyle^{23,24,27,28,30-34}, managing symptoms^{26,29,32}, talking with friends and family³¹⁻³³, problem solving^{25,29,34}, working with the care team^{31,32}, participating in social activities^{27,33}, managing medications³¹, and community resources²⁷. The majority of conversational agents were chatbots^{23-31,33} and almost half used a frame-based system^{26,28-30,32} to keep track of information the user provided and information the system still needed.

Most conversational agents had a mixed dialogue initiative^{25,26,28-30,33}, thereby, allowing the user or system to lead the conversation. The majority of agents had a mobile app interface, which included custom apps or SMS platforms such as Telegram or WeChat^{26-28,30,32,33}. Nearly all were based on theories for behavior change or contained evidence-based content^{24-30,32-34}. Across studies, cognitive behavioral therapy was most commonly used to guide the language the agent used and the direction of the conversation based on the user's utterances. Few described design principles used to develop the content or types of interactions, with only two studies reporting the use of human-centered design³¹ or the Integrate, Design, Assess, and Share (IDEAS) framework^{27,35}.

Table 2. Conversational agent attributes.

Study	Chronic Disease or Population	Type	Theoretical Framework or Evidence Base	Self-Management Skills
Baskar et al, 2015 ²³	Older Adults	Chatbot (finite-state)	N/A	<ul style="list-style-type: none"> • Maintaining a healthy lifestyle
Elmasri et al, 2016 ²⁴	Substance Use	Chatbot (finite-state)	Alcohol Use Disorders Identification Test	<ul style="list-style-type: none"> • Maintaining a healthy lifestyle
Fitzpatrick et al, 2017 ²⁶	Depression	Chatbot (frame-based)	Cognitive behavior therapy	<ul style="list-style-type: none"> • Managing symptoms
Gaffney et al, 2013 ²⁵	Depression	Chatbot (finite-state)	Perceptual control theory	<ul style="list-style-type: none"> • Problem solving
Kazemi et al, 2014 ^{27*}	Substance Use	Chatbot**	Ecological momentary interventions, motivational interviewing, transtheoretical model of change	<ul style="list-style-type: none"> • Maintaining a healthy lifestyle • Community resources • Participating in social activities
Ly et al, 2017 ²⁸	Depression	Chatbot (frame-based)	Cognitive behavior therapy	<ul style="list-style-type: none"> • Maintaining a healthy lifestyle
Schroeder et al, 2018 ²⁹	Depression	Chatbot (frame-based)	Dialectical behavior therapy	<ul style="list-style-type: none"> • Managing symptoms • Problem solving
Stein et al, 2017 ³⁰	Diabetes	Chatbot (frame-based)	Cognitive behavior therapy, Diabetes Prevention Program	<ul style="list-style-type: none"> • Maintaining a healthy lifestyle
Tsiourti et al, 2014 ³¹	Older Adults	Chatbot**	N/A	<ul style="list-style-type: none"> • Maintaining a healthy lifestyle • Working with the care team • Talking with friends and family • Managing medications
van Heerden et al, 2017 ³²	HIV/AIDS	Task-oriented (frame-based)	CDC guidelines for HIV counseling in non-clinical setting	<ul style="list-style-type: none"> • Maintaining a healthy lifestyle • Managing symptoms • Working with the care team • Talking with friends and family
Wang et al, 2018 ³³	Substance Use	Chatbot (finite-state)	PubMed medical information retrieval dataset	<ul style="list-style-type: none"> • Maintaining a healthy lifestyle • Talking with friends and family • Participating in social activities
Watson et al, 2012 ³⁴	Overweight or Obesity	Task-oriented (finite-state)	Behavioral and social cognitive theory	<ul style="list-style-type: none"> • Maintaining a healthy lifestyle • Problem solving

*Kazemi et al, 2018 was used to extract conversational agent attributes³⁵.

**Dialogue management system was N/A.

Risk of Bias. Risk of bias of one study was rated as good, five as fair, and six as poor. There was selection bias across studies with the majority using a convenience sample or not adequately describing participant characteristics. Of the six non-laboratory based experimental studies, participant completion ranged from 56 to 96%, and only three studies conducted an intent-to-treat analysis^{26,28,34}. Measurement bias included self-reported measurements and questionnaires that were not valid or reliable, which may be due to the lack of established instruments to measure the quality of interactions with a conversational agent. Many studies did not assess potential confounders such as the participant's duration of the condition, comorbidities, health literacy, motivation, and use of technology.

Discussion

Principle Findings. While limited, early evidence suggests that conversational agents are acceptable, usable, and may be effective in supporting patients in self-management of chronic diseases particularly for mental health conditions such as depression or substance use. Outcomes from reviewed studies were mainly evaluated on usability, usage, and self-reported clinical measurements. Overall, participants reported a positive attitude and moderate to high satisfaction with agents, but there were concerns for privacy, the potential to reduce independence, and for having repetitive or shallow content. There were improvements in several patient-reported outcome scores between the conversational agent intervention and control groups in several studies^{26,28,29}. However, the lack of methodological rigor and heterogeneity in study design make it challenging to interpret results within and across studies.

In this systematic review, we extended prior reviews by assessing the theoretical frameworks, features, content, and design principles of text-based conversational agents for self-management in those with chronic diseases. There was diverse content across the conversational agents. In contrast to the mHealth literature^{9,10}, we found that the majority of conversational agents were based upon theoretical grounding. Agents primarily leveraged the self-management skill of maintaining a healthy lifestyle, but few provided information about community resources or assistance with managing medications. Similar to digital health solutions, conversational agents vary considerably, and it is unclear how their effectiveness may differ based on sociodemographic and clinical characteristics as these were not captured across all studies. Not surprisingly, very few studies reported using established design principles, such as participatory or user-centered design, and none used heuristic evaluation methods.

Comparison to Prior Work. To our knowledge, there have been no systematic reviews on the use of conversational agents for chronic disease self-management, but other health-related reviews have been conducted with similar findings³⁶⁻⁴¹. We found that the existing literature primarily contains small scale studies that focus on mental health conditions³⁶⁻⁴⁰. Many of these studies are in the development or pilot phase and do not contain samples that are generalizable across other populations. Study participants from research included in this review were predominantly young, female, Caucasian, and owned a smartphone, which is not necessarily representative of U.S. patients with chronic diseases. Excluding populations that may benefit the most from using these technologies, such as older adults or those with physical limitations, from design and usability studies could widen disparities. Despite our focus on chronic diseases, the longest study duration was only four months³⁰, and no agents accounted for time since diagnosis to allow for more complex or nuanced interactions. Providing improved continuity and information to users comes with a tradeoff, as agent-based systems are computationally intensive and require more sophisticated natural language capabilities, large training datasets, and a deeper semantic representation. Ideally, sophisticated dialogue and user models that prioritize safety and efficacy and can handle natural language inputs should take into account the users' goals, intentions, dialogue history, and context.

In accordance with prior reviews, we also found that studies lacked a comprehensive understanding of patients' needs^{37,38}. Patients may require varying levels of support depending on sociodemographic characteristics, health status, and cultural factors, and these needs largely remain implicit and unaddressed in the existing body of research^{36,39}. For example, patients with low intrinsic motivation may benefit more from human support than those motivated to work independently, and understanding user motivation could contribute to the optimal timing and type of support provided by conversational agents³⁹. While there were no studies that explicitly assessed patient motivation, two studies measured levels of engagement or experience with technology at baseline^{25,31}. These findings suggest the timing and nature of support and how individual characteristics affect one's ability to self-manage health and use conversational technologies should be design considerations.

Implications for Healthcare and Research. The use of conversational agents for chronic disease self-management also has a number of potential applications for clinical care and research. In clinical settings, conversational agents could facilitate social, emotional, relational, and task support that could connect users to health professionals, resources, information, and peers. However, safety protocols are needed prior to leveraging conversational agents in clinical care settings, particularly for agents with unconstrained natural language input capabilities⁴². For example, if

the system does not accurately understand the user's input, it could provide inappropriate medical information which may result in an adverse outcome. Confirming user input, having a default fallback intent for out-of-domain utterances, or constraining user input to menu choices may mitigate some patient safety concerns. These tradeoffs must be considered for user experience, but balanced against the relative risks to patient safety.

The field of health conversational agents, which lies at the intersection of computational linguistics and health informatics⁴³, will likely accelerate with the adoption of conversational developer platforms, such as Amazon Lex⁴⁴, Facebook Wit.ai⁴⁵, Google Dialogflow⁴⁶, and Microsoft Bot Framework⁴⁷. As mobile phones become increasingly integrated into people's lives, conversational agents will likely be deployable on new and existing mobile apps, such as social media, messaging platforms, or even other emerging technologies. However, many of these platforms do not meet the requirements of the Health Insurance Portability and Accountability Act (HIPAA) to protect health-related data, which is a potential limitation of implementing conversational agents for use in clinical care. Currently, Amazon Alexa is the only platform that provides a HIPAA-compliant environment to build apps that transmit and receive protected health information, though it is only available to select developers at this time⁴⁸. Further, the Food and Drug Administration has not made an explicit statement regarding the level of enforcement for health-related conversational agents. It is likely that conversational agents would be regulated like mobile apps, where mobile app entities are not considered medical device manufacturers unless the app delivers care or makes care decisions⁴⁹. In addition to regulatory and legal provisions, reimbursement mechanisms and the ability to incorporate actionable information from agents into the clinical workflow remains unexplored. Ethical concerns related to the use of assistive technologies, such as the quality of information, over-reliance, and potentially further exacerbating health disparities if user-centered design principles are not utilized, are also barriers to using conversational agents to support self-management.

Limitations. This review has several limitations. We did not include voice-based conversational agents because they offer a different user experience than text-based agents¹⁵. Also, we only included the features, content, theoretical frameworks, and design principles of the conversational agents that were explicitly stated in the manuscripts, so it is possible that some of the characteristics may not be reflected in our review. We rated the majority of papers as poor or fair quality due to the lack of rigor in many of the studies, which suggests the conclusions may not be generalizable given the high degree of bias. Due to the heterogeneity of study designs, we could not assess the strength of evidence in a meta-analysis.

Conclusions

Given the growing burden of chronic diseases in the U.S., text-based conversational agents to support self-management could assist patients to move beyond passively consuming information to actively engaging in disease management. Currently, conversational agents seem to have elementary dialogue management systems that do not take into account the users' preferences, goals, or history, and the design may need to evolve to better meet user needs. The rich linguistic data generated from agents could also provide additional insights into the patient's emotional or physical state, could facilitate decision making and self-management for the patient, and provide valuable information for the care team. At this early stage, there are only a few conversational agents that are targeted towards chronic diseases, and most are focused on depression and self-management through maintaining a healthy lifestyle. Future research should assess the influence of more sophisticated dialogue and interactions and leverage established user-centered design principles to tailor agents to support self-management of chronic diseases. Additional investigation is needed to rigorously assess the characteristics of agents that may be most useful for self-management based on the user's motivation and context, health status, and psychosocial attributes. Safety, privacy, ethical, and regulatory issues should also be addressed as conversational agents are implemented into real-world care settings.

Acknowledgements

AG is supported by the National Institutes of Health National Library of Medicine training grant (5T15LM012500-03). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. Buttorff C, Ruder T, Bauman M. Multiple chronic conditions in the United States [Internet]. 2017 [cited 22 May 2019]. Available from: <https://www.rand.org/pubs/tools/TL221.html>
2. Gerteis J, Izrael D, Deitz D, LeRoy L, Ricciardi R, Miller T, et al. Multiple chronic conditions chartbook. Agency for Healthcare Research and Quality [Internet]. 2014 [cited 1 May 2019]. Available from: <https://www.ahrq.gov/professionals/prevention-chronic-care/decision/mcc/resources.html>

3. Centers for Medicare and Medicaid Services. National health expenditures 2010: sponsor highlights [Internet]. 2010 [cited 22 May 2019]. Available from: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/>
4. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. *Patient Educ Couns* 2002;48(2):177-87.
5. Pew Research Center. Mobile fact sheet [Internet]. 2019 [cited 22 May 2019]. Available from: <http://www.pewinternet.org/fact-sheet/mobile/>
6. Thakkar J, Kurup R, Laba TL, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. *JAMA Intern Med* 2016;176(3):340-9.
7. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. *Cochrane Database Syst Rev* 2016;4:CD006611.
8. Williams AD. Use of a text messaging program to promote adherence to daily physical activity guidelines: a review of the literature. *Bariatric Nursing and Surgical Patient Care* 2012;7(1).
9. Arambepola C, Ricci-Cabello I, Manikavasagam P, Roberts N, French DP, Farmer A. The impact of automated brief messages promoting lifestyle changes delivered via mobile devices to people with type 2 diabetes: a systematic literature review and meta-analysis of controlled trials. *J Med Internet Res* 2016;18(4):e86.
10. Alessa T, Abdi S, Hawley MS, de Witte L. Mobile apps to support the self-management of hypertension: systematic review of effectiveness, usability, and user satisfaction. *JMIR Mhealth Uhealth* 2018;6(7):e10723.
11. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. *J Med Internet Res* 2015;17(2):e52.
12. Jurafsky D, Martin J. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd ed. Upper Saddle River, NJ: Pearson Prentice Hall 2009.
13. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Educ Couns* 2005;59(1):21-30.
14. Klein J, Moon Y, Picard RW. This computer responds to user frustration: theory, design, and results. *Interacting with Computers* 2002;14(2):119-140.
15. Guy I. Searching by talking: analysis of voice queries on mobile web search. *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*; 2016:35-44.
16. Kappas A, Kuster D, Basedow C, Dente P. A validation study of the affectiva Q-Sensor in different social laboratory situations. *53rd Annual Meeting of the Society for Psychophysiological Research*; 2013.
17. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700.
18. Lorig K, Holman H, Sobel D, Laurent D, Gonzalez V, Minor M. *Living a Healthy Life with Chronic Conditions: For Ongoing Physical and Mental Health Conditions*. 4th ed. Boulder, CO: Bull Publishing Company 2013.
19. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. *Ann Behav Med* 2003;26(1):1-7.
20. Centers for Medicare and Medicaid Services. Chronic conditions [Internet]. 2019 [cited 2019 May 22]. Available from: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/CC_Main.html
21. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37-46.
22. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Viera A, Crotty K, et al. Health literacy interventions and outcomes: an updated systematic review. *Evidence Report/Technology Assessment No. 199*. AHRQ Publication Number 11-E006, 2011.
23. Baskar J, Lindgren H. Human-agent dialogues on health topics - an evaluation study. *Communications in Computer and Information Science* 2015;524:28-39.
24. Elmasri D, Maeder A. A conversational agent for an online mental health intervention. *Brain Informatics and Health: International Conference* 2016;243-251.
25. Gaffney H, Mansell W, Edwards R, Wright J. Manage Your Life Online (MYLO): a pilot trial of a conversational computer-based intervention for problem solving in a student sample. *Behav Cogn Psychother* 2014;42(6):731-46.
26. Fitzpatrick K, Darcy A, 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(2):e19.

27. Kazemi DM, Cochran AR, Kelly JF, Cornelius JB, Belk C. Integrating mHealth mobile applications to reduce high risk drinking among underage students. *Health Education Journal* 2013;73(3).
28. Ly KH, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interv* 2017; 10:39-46.
29. Schroeder J, Wilkes C, Rowan K, Toledo A, Paradiso A, Czerwinski M, et al. Pocket skills: a conversational mobile web app to support dialectical behavioral therapy. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* 2018;1-15.
30. Stein N, Brooks K. A fully automated conversational artificial intelligence for weight loss: longitudinal observational study among overweight and obese adults. *JMIR Diabetes* 2017;2(2):e28.
31. Tsiourti C, Joly E, Wings C, Moussa MB, Wac K. Virtual assistive companions for older adults: qualitative field study and design implications. *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare* 2014;57-64.
32. van Heerden A, Ntinga X, Vilakazi K. The potential of conversational agents to provide a rapid HIV counseling and testing services. *International Conference on the Frontiers and Advances in Data Science* 2017.
33. Wang H, Zhang Q, Ip M, Lau JT. Social media-based conversational agents for health management and interventions. *Computer* 2018;51(8):26-33.
34. Watson A, Bickmore T, Cange A, Kulshreshtha A, Kvedar J. An internet-based virtual coach to promote physical activity adherence in overweight adults: randomized controlled trial. *J Med Internet Res* 2012;14(1):e1.
35. Kazemi DM, Borsari B, Levine MJ, Lamberson KA, Dooley B. REMIT: Development of a mHealth theory-based intervention to decrease heavy episodic drinking among college students. *Addiction Research and Theory* 2018;26(5):377-385.
36. Scholten MR, Kelders SM, van Gemert-Pijnen JE. Self-guided web-based interventions: scoping review on user needs and the potential of embodied conversational agents to address them. *J Med Internet Res* 2017;19(11):e383.
37. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc* 2018;25(9):1248-1258.
38. Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. *J Med Internet Res* 2017;19(8):e267.
39. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. *J Med Internet Res* 2017;19(5):e151.
40. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *Can J Psychiatry* 2019;64(7):456-464.
41. Kramer LL, ter Stal S, Mulder BC, de Vet E, van Velsen L. Developing embodied conversational agents for coaching people in a healthy lifestyle: scoping review. *J Med Internet Res* 2020;22(2):e14058.
42. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and consumer safety risks when using conversational assistants for medical information: an observational study of Siri, Alexa, and Google Assistant. *J Med Internet Res* 2018;20(9):e11510.
43. Bickmore T, Giorgino T. Health dialog systems for patients and consumers. *Journal of Biomedical Informatics* 2006;39(5):556-571.
44. Amazon. Lex [Internet]. 2020 [cited 2020 March 3]. Available from: <https://aws.amazon.com/lex/>
45. Wit.ai, Inc. Natural language for developers [Internet]. 2020 [cited 2020 March 3]. Available from: <https://wit.ai/>
46. Google. Dialogflow [Internet]. 2020 [cited 2020 March 3]. Available from: <https://dialogflow.com/>
47. Microsoft. Bot framework [Internet]. 2020 [cited 2020 March 3]. Available from: <https://dev.botframework.com/>
48. Amazon. Introducing new Alexa healthcare skills [Internet]. 2020 [cited 2020 March 3]. Available from: <https://developer.amazon.com/blogs/alexa/post/ff33dbc7-6cf5-4db8-b203-99144a251a21/introducing-new-alexa-healthcare-skills>
49. Food and Drug Administration (US). Mobile medical applications [Internet]. 2018 [cited 2019 May 22]. Available from: <https://www.fda.gov/MedicalDevices/DigitalHealth/MobileMedicalApplications/ucm255978.htm>