Voice-based Conversational Agents for the Prevention and Management of Chronic and Mental Conditions: A Systematic Literature Review

Caterina Bérubé, Theresa Schachner, Roman Keller, Elgar Fleisch, Florian v. Wangenheim, Filipe Barata, Tobias Kowatsch

Submitted to: Journal of Medical Internet Research
on: November 30, 2020

Table of Contents

Voice-based Conversational Agents for the Prevention and Management of Chronic and Mental Conditions: A Systematic Literature Review

Caterina Bérubé[1] MSc; Theresa Schachner[1] MSc; Roman Keller[2] MSc; Elgar Fleisch[1, 2, 3] Prof Dr; Florian v. Wangenheim[1, 2] Prof Dr; Filipe Barata[1] PhD; Tobias Kowatsch[1, 2, 3] PhD

[1]Department of Management, Technology, and Economics (D-MTEC) ETH Zurich Zurich CH
[2]Future Health Technologies (FHT) Campus for Research Excellence and Technological Enterprise (CREATE) Singapore-ETH Centre Singapore SG
[3]Institute of Technology Management (ITEM) University of St. Gallen St. Gallen CH

Corresponding Author:
Caterina Bérubé MSc
Department of Management, Technology, and Economics (D-MTEC)
ETH Zurich
WEV G 214
Weinbergstrasse 56/58
Zurich
CH

Abstract

Background: Chronic and mental conditions are increasingly prevalent worldwide. As devices in our everyday lives offer more and more voice-based self-service, voice-based conversational agents (VCAs) have the potential to support the prevention and management of these conditions in a scalable way. VCAs allow for a more natural interaction compared to text-based conversational agents, facilitate input for users who cannot type, allow for routine monitoring and support when in-person healthcare is not possible, and open the doors to voice and speech analysis. The state of the art of VCAs for chronic and mental conditions is, however, unclear.

Objective: This systematic literature review aims to provide a better understanding of state-of-the-art research on VCAs delivering interventions for the prevention and management of chronic and mental conditions.

Methods: We conducted a systematic literature review using PubMed Medline, EMBASE, PsycINFO, Scopus, and Web of Science databases. We included primary research that involved the prevention or management of chronic or mental conditions, where the voice was the primary interaction modality of the conversational agent, and where an empirical evaluation of the system in terms of system accuracy and/or in terms of technology acceptance was included. Two independent reviewers conducted screening and data extraction and measured their agreement with Cohen's kappa. A narrative approach was applied to synthesize the selected records.

Results: Twelve out of 7'170 articles met the inclusion criteria. The majority of the studies (N=10) were non-experimental, while the remainder (N=2) were quasi-experimental. The VCAs provided behavioral support (N=5), a health monitoring service (N=3), or both (N=4). The VCA services were delivered via smartphone (N=5), tablet (N=2), or smart speakers (N=3). In two cases, no device was specified. Three VCAs targeted cancer, while two VCAs each targeted diabetes and heart failure. The other VCAs targeted hearing-impairment, asthma, Parkinson's disease, dementia and autism, "intellectual disability", and depression. The majority of the studies (N=7) assessed technology acceptance but only a minority (N=3) used validated instruments. Half of the studies (N=6) reported either performance measures on speech recognition or on the ability of VCA's to respond to health-related queries. Only a minority of the studies (N=2) reported behavioral measure or a measure of attitudes towards intervention-related health behavior. Moreover, only a minority of studies (N=4) reported controlling for participant's previous experience with technology.

Conclusions: Considering the heterogeneity of the methods and the limited number of studies identified, it seems that research on VCAs for chronic and mental conditions is still in its infancy. Although results in system accuracy and technology acceptance are encouraging, there still is a need to establish evidence on the efficacy of VCAs for the prevention and management of chronic and mental conditions, both in absolute terms and in comparison to standard healthcare.

(JMIR Preprints 30/11/2020:25933)
DOI: https://doi.org/10.2196/preprints.25933

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

- ✓ **Please make my preprint PDF available to anyone at any time (recommended).**
- Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
- Only make the preprint title and abstract visible.
- No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

- ✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
- Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
- Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Voice-based Conversational Agents for the Prevention and Management of Chronic and Mental Conditions: A Systematic Literature Review

Caterina Bérubé[1], MSc; Theresa Schachner[1]*, MSc; Roman Keller[2]*, MSc; Elgar Fleisch[1,2,3,], Prof. Dr., Florian von Wangenheim[1,2], Prof Dr; Filipe Barata[1†], PhD; Tobias Kowatsch[1,2,3†], PhD
[1]Department of Management, Technology, and Economics (D-MTEC), ETH Zurich, Zurich, Switzerland
[2]Future Health Technologies (FHT), Campus for Research Excellence and Technological Enterprise (CREATE), Singapore-ETH Centre, Singapore, Singapore
[3]Institute of Technology Management (ITEM), University of St. Gallen, St. Gallen, Switzerland

* shared second authorship
† shared last authorship

Abstract

Background: Chronic and mental conditions are increasingly prevalent worldwide. As devices in our everyday lives offer more and more voice-based self-service, voice-based conversational agents (VCAs) have the potential to support the prevention and management of these conditions in a scalable way. VCAs allow for a more natural interaction compared to text-based conversational agents, facilitate input for users who cannot type, allow for routine monitoring and support when in-person healthcare is not possible, and open the doors to voice and speech analysis. The state of the art of VCAs for chronic and mental conditions is, however, unclear.
Objective: This systematic literature review aims to provide a better understanding of state-of-the-art research on VCAs delivering interventions for the prevention and management of chronic and mental conditions.
Methods: We conducted a systematic literature review using PubMed Medline, EMBASE, PsycINFO, Scopus, and Web of Science databases. We included primary research that involved the prevention or management of chronic or mental conditions, where the voice was the primary interaction modality of the conversational agent, and where an empirical evaluation of the system in terms of system accuracy and/or in terms of technology acceptance was included. Two independent reviewers conducted screening and data extraction and measured their agreement with Cohen's kappa. A narrative approach was applied to synthesize the selected records.
Results: Twelve out of 7'170 articles met the inclusion criteria. All studies (N=12) were non-experimental. The VCAs provided behavioral support (N=5), a health monitoring service (N=3), or both (N=4). The VCA services were delivered via smartphone (N=5), tablet (N=2), or smart speakers (N=3). In two cases, no device was specified. Three VCAs targeted cancer, while two VCAs each targeted diabetes and heart failure. The other VCAs targeted hearing-impairment, asthma, Parkinson's disease, dementia and autism, "intellectual disability", and depression. The majority of the studies (N=7) assessed technology acceptance but only a minority (N=3) used validated instruments. Half of the studies (N=6) reported either performance measures on speech recognition or on the ability of VCA's to respond to health-related queries. Only a minority of the studies (N=2) reported behavioral measure or a measure of attitudes towards intervention-related health behavior. Moreover, only a minority of studies (N=4) reported controlling for participant's previous experience with technology.
Conclusions: Considering the heterogeneity of the methods and the limited number of studies

identified, it seems that research on VCAs for chronic and mental conditions is still in its infancy. Although results in system accuracy and technology acceptance are encouraging, there still is a need to establish evidence on the efficacy of VCAs for the prevention and management of chronic and mental conditions, both in absolute terms and in comparison to standard healthcare.

Keywords: Voice; Speech; Delivery of Health Care; Noncommunicable Diseases; Conversational agents; Monitoring; Support; Chronic Disease; Mental Health; Systematic Literature Review

Introduction

Chronic and mental conditions are increasingly prevalent worldwide. According to the World Health Statistics of 2020, non-communicable diseases (e.g., cardiovascular diseases, cancer, chronic respiratory diseases, and diabetes) and suicide are still predominant causes of death in 2016 [1, 2]. Although the underlying causes of these conditions are complex, behavior remains an important factor for their prevention and management. As the healthcare system is currently unfit to sustain the prevention and management of chronic and mental conditions while containing its costs, continuous and personalized smartphone-based interventions have been developed to provide scaled-up behavioral support [3-6]. On the same note, conversational agents have been proven a valuable tool to deliver digital health interventions [7-9]. In particular, voice-based conversational agents (VCA) have been shown to provide high user satisfaction in delivering interventions to influence healthy lifestyles [6].

VCAs can recognize human speech and in turn respond with synthesized speech. The human input is converted to an intent, triggering a specific information retrieval or a specific function. This modality of interaction allows for hands-free access to some basic functions, such as searching for information on the internet, managing calendars, playing media content, calling, texting, and emails, as well as controlling internet-of-things devices or telling jokes [10, 11]. Just as text-based [12, 13] and embodied [14] conversational agents, VCAs have the potential to form an “alliance” [15] or “rapport” [16] with the patient through conversation, which is beneficial to treatment outcomes [17, 18]. However, compared to textual and visual, voice-based interaction has several advantages. First, it leverages the naturalness [19, 20] and social presence [21, 22] of human-to-human conversation. Second, it facilitates input for users with low literacy, or with visual [23], intellectual [24], motor, linguistic, and cognitive disabilities [25], and can support more natural health routine tasks when in-person health care is not possible [26]. Third, it opens the doors to voice or speech analysis, whereas features of the patient’s utterances can be passively monitored to derive health states [27-29]. Given the lack of agreement on the terminology [6], we will refer to VCAs to indicate the broad technology of dialogue systems interacting with humans through speech recognition and synthesis.

VCAs are currently available on 2.5 billion devices worldwide, with smartphones being the leading type of device, followed by smart speakers, computers. They can be found even in wearable technology, cars, and appliances [30, 31]. Moreover, numerous health-related applications for VCAs are already available [32]. Thus, these systems are more and more used in our daily life and able to assist in the healthcare domain. In particular, commercial VCAs like Alexa, Google Assistant, and Siri are increasingly adopted and health-related products that are compatible with their technology are being developed [33-37]. Although there is still room for improvement [38], curiosity in using VCAs for health care is growing. VCAs are used to retrieve health-related information (e.g., symptoms, medication, nutrition, and healthcare facilities) [30, 39]. This interest is even stronger in low-income households (i.e. < $50,000 a year). Furthermore, when considering the accessibility of the voice modality for users with low literacy, VCAs could facilitate health management in countries, where the education index is still relatively low [40] and smartphones are increasingly penetrating daily life [41] (e.g., Brazil, Indonesia, Kenya, Mexico, Philippines, or South Africa).

To the best of our knowledge, only one scoping review focused on VCAs for healthy lifestyle behaviors [6]. The authors included research promoting self-management skills and healthy lifestyle behaviors in general and found that, although showing the feasibility of VCAs for health, the evidence was mostly preliminary. In contrast, our contribution lies in a systematic review of VCA applications dedicated to the prevention and management of chronic and mental conditions. Also, to have a broader overview of the current state of research, we include evidence from both journals and conference papers and provide an overview of aspects affecting technology adoptions, i.e., system and user performance, ease of use, and attitude towards the target health behavior [42].

This systematic literature review aims to provide a better understanding of state-of-the-art research on conversational agents delivering health interventions through voice-based interaction and to deliver an overview of the methods and evaluations performed. We focus on VCAs specifically dedicated to the prevention and management of chronic and mental conditions. Hence, in this paper, we seek to answer the following two questions:

1. What is the current state-of-the-art of evidence in favor of VCAs for the prevention and management of chronic and mental conditions?
2. What are the methods used to evaluate them?

Methods

Reporting standards

This systematic review is compliant with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist [43] (see Multimedia Appendix 1 for an overview of the study protocol).

Search strategy

We conducted a systematic search of the literature available in July 2020, using electronic databases PubMed Medline, EMBASE, PsycINFO, Scopus, and Web of Science. These databases were chosen as they cover relevant aspects in the fields of medicine, technology, and interdisciplinary research and have also been used in other systematic reviews covering similar topics [7, 8].
Search terms included items describing the constructs “voice modality”, “conversational agent” and “health” (see Multimedia Appendix 2 for an overview of the search strategy).

Selection criteria

We included studies if they (1) were primary research studies involving the prevention, treatment, or management of health conditions related to chronic diseases or mental disorders in patients, (2) involved a conversational agent, (3) the agent used voice as main interaction modality, and (4) the study included an empirical evaluation of the system in terms of system accuracy (e.g., speech recognition, quality of answers) and/or in terms of technology acceptance (e.g., user experience, usability, likability, engagement).
Articles were excluded if they (1) involved any form of animation or visual representation, e.g., embodied agents, virtual humans, or robots, (2) involved any form of healthcare service via telephone (e.g., interactive voice response), (3) focused on testing a machine learning algorithm, (4) did not target a specific patient population, chronic [44] or mental [45] condition.
We also excluded non-English papers, workshop papers, literature reviews, posters, PowerPoint

presentations, articles presented at doctoral colloquia, or if the full text was not accessible for the study authors.

Selection process

We downloaded all references and inserted them into an Excel spreadsheet (Microsoft Corporation), and removed duplicates. Two independent investigators conducted the screening for inclusion and exclusion criteria in three phases: first, we assessed the records' titles, then their abstracts, and, finally, the full texts. After each of these phases, we calculated Cohen kappa to measure inter-rater agreement between the two investigators. The interpretation of the Cohen's kappa coefficient was based on the categories developed by Douglas Altman: 0.00-0.20 (poor), 0.21-0.40 (fair), 0.41-0.60 (moderate), 0.61-0.80 (good), and 0.81-1.00 (very good) [46, 47]. The two raters consulted a third investigator in case of disagreements.

Data Extraction

Two investigators extracted data from the eligible articles into an Excel spreadsheet with 52 columns containing information on the following aspects: (1) general information about the included papers, (2) voice-based interaction, (3) conversational agents, (4) targeted health conditions, (5) participants, (6) design, (7) measures, (8) main findings, and (9) additional study information such as funding information or conflicts of interest (see Multimedia Appendix 3 for a complete overview of the study characteristics).

We chose a narrative synthesis of the results, and discussed and resolved any inconsistencies in the individual data extractions with a third investigator.

Risk of Methodological Bias

The choice of an appropriate risk of bias assessment tool was arbitrary, given the prevalence of conference papers and a wide variety of research designs in the included studies. Nevertheless, we wanted to evaluate the selected research concerning the transparency of reporting and the quality of the evidence. After extensive team discussions, the investigators decided to follow the approach of Maher et al [48], who devised a risk-of-bias assessment tool based upon the Consolidated Standards of Reporting Trials (CONSORT) checklist[49]. The tool comprises 25 items and assigns scores of 0 or 1 to each item, indicating if the respective study satisfactorily met the criteria. Higher total scores indicate a smaller risk of methodological bias. As the CONSORT checklist was originally developed for controlled trials and no such trials were included in our set of studies, we decided to exclude and adapt certain items as they were considered out of scope for this type of study. We excluded 3.b ("Trial design"), 6.b ("Outcomes"), 7.b ("Sample size"), 12.b ("Statistical methods"), 14.b ("Recruitment"). Finally, item 17.b ("Outcomes and estimation") was excluded and 17.a was separated into two sub-criteria (i.e., "a. provides the estimated effect size", "b. provides precision"). Two investigators conducted the risk-of-bias assessment independently and the differences were resolved in a consensus agreement (see details in Multimedia Appendix 4).

Results

Selection and Inclusion of Studies

In total, we screened 7'170 de-duplicated citations from electronic databases (see Figure 1). Out of these, we excluded 6'910 papers during title screening. We further excluded 140 papers in the abstract-screening process, which left us with 120 papers for full-text screening. After assessing the full texts, we found 108 not to be qualified (see Figure 1 for an overview of the reasons for exclusion and the number of excluded records) and considered 12 papers as qualified for inclusion and analysis (see Table 1).

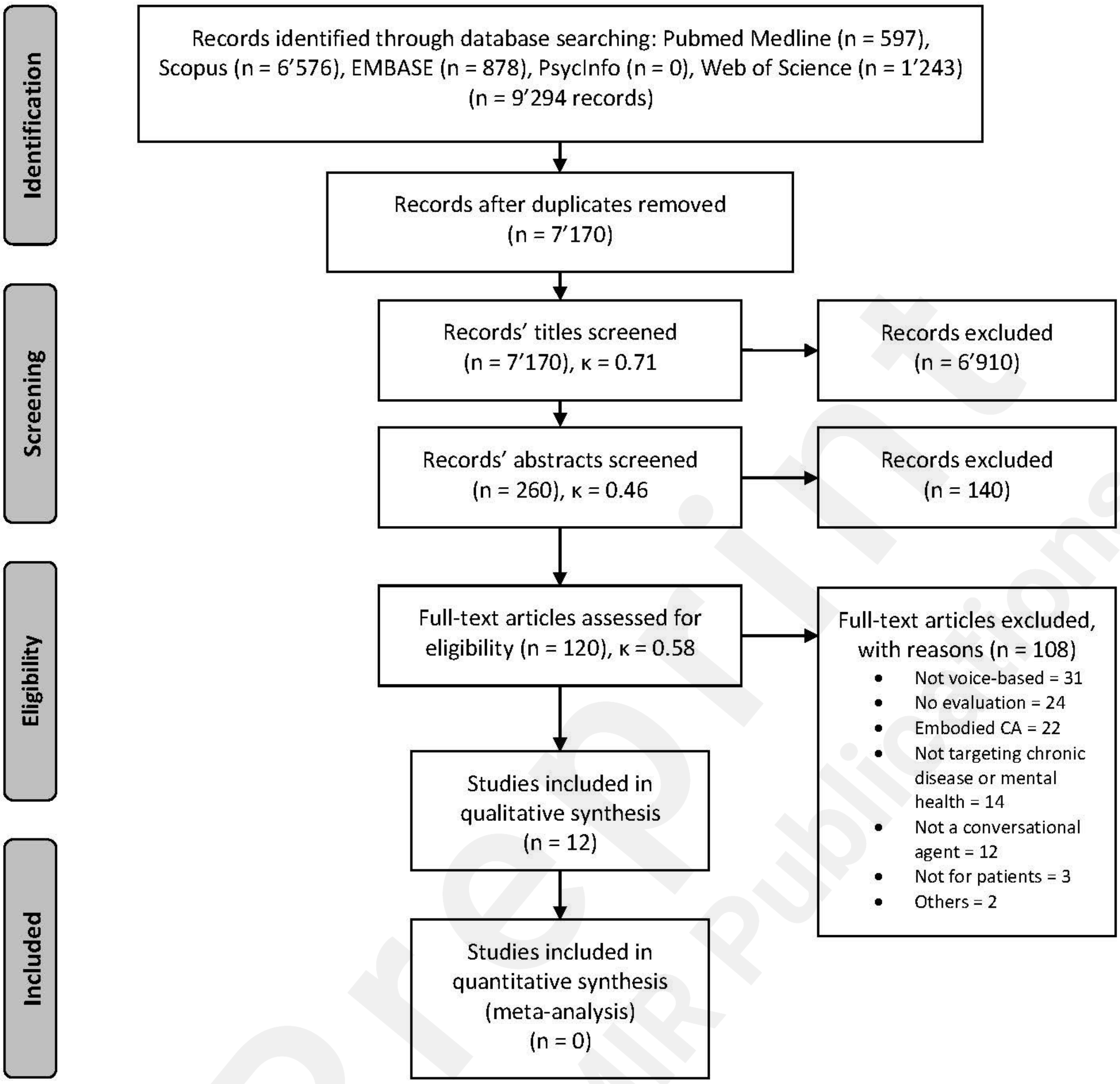


Figure 1. PRISMA flow diagram of included studies.

Table 1. Overview and characteristics of included records.[a]

Study ID, Publication Year	Study Aim	Type of study participants, Number of study participants	Addressed Medical Condition	Voice-enabled Device Type	Intervention Category
Amith et al., 2019 [50]	Development and acceptance evaluation	Healthy adults with at least one child under the age of 18 (N=16)	Cancers associated with HPV	Tablet	Support
Amith et al., 2020 [51]	Development and acceptance evaluation	Healthy young adults between 18 and 16 years old (N=24)	Cancers associated with HPV	Tablet	Support
Boyd and Wilson, 2018	Criteria-based performance	Authors as raters (N=2)	Cancers associated	Smartphone	Support

[52]	evaluation of commercial conversational agent		with smoking		
Cheng et al., 2019 [53]	Development and acceptance evaluation	Elderly (N=10)	Diabetes (T2)	Smart speaker	Monitoring, Support
Galescu et al., 2009 [54]	Development and performance evaluation	Chronic heart failure patients (N=14)	Heart Failure	Not specified	Monitoring
Greuter and Balandin, 2019 [55]	Development and performance evaluation	Adults with lifelong intellectual disability (N=9)	Intellectual disability	Smart speaker	Support
Ireland et al., 2016 [56]	Development and acceptance evaluation	Adults recruited on campus (N=33)	Parkinson Disease, Dementia, Autism	Smartphone	Monitoring
Kadariya et al., 2019 [57]	Development and acceptance evaluation	Clinicians and Researchers (N=16)	Asthma	Smartphone	Monitoring, Support
Lobo et al., 2017 [58]	Development and acceptance evaluation	Healthy adults working regularly with senior patients (N=11)	Heart Failure	Smartphone	Monitoring, Support
Ooster et al., 2019 [59]	Development and performance evaluation	Normal-hearing (N=6)	Hearing-impairment	Smart speaker	Monitoring
Rehman et al., 2020 [60]	Development and performance & acceptance evaluation	Adults affiliated with the university (N=33)	Diabetes (T1, T2, gestational) and Glaucoma	Smartphone	Monitoring, Support
Reis et al., 2018 [61]	Criteria-based performance evaluation of commercial conversational agent	Not specified (N=Not specified)	Depression	Not specified	Support

[a]Abbreviations: HPV = Human Papillomavirus, T1= Type 1, T2=Type 2.

Characteristics of the included studies

The publication years of the selected records ranged between 2009 and 2020, whereas the majority (N=5) was published in 2019. Seven of the selected records were conference papers and five were journal articles.

The majority (N=10) of the selected papers, developed and evaluated VCA [50, 51, 53-60], while two [52, 61] aimed at reporting a criteria-based performance evaluation of existing commercial conversational agents (e.g., Google Assistant, Siri). Among the papers developing and evaluating a VCA, six [50, 51, 53, 56-58] assessed the technology acceptance of the VCA, while three [54, 55, 59] assessed the system accuracy. Only one [60] assessed both performance and acceptance evaluation.

All studies (N=12) were non-experimental [50-61], i.e., they did not include any experimental manipulation. Four [52, 57, 59, 61] did not explicitly specify what study design they used, while the others provided a label. One study stated conducting a feasibility evaluation [54], one a focus group study [56], one a qualitative assessment of effectiveness and satisfaction [53], and one a case study [60]. Furthermore, one conducted a pilot study [55], two declared deploying a Wizard-of-Oz (WOz) experiment [50, 51], and one a usability study [58].

An overview of the included studies can be found in (Table 1; see all details in Multimedia Appendix 3).

Main findings

System accuracy

Half (N=6) of the studies [52, 54, 55, 59-61] evaluated the system's accuracy. Four of those [54, 55, 59, 60] described precise speech recognition performance, whereas three [54, 59, 60] reported good or very good speech recognition performance, and one [55] found mediocre recognition accuracy, with single letter responses being slightly better recognized than word-based responses (see Multimedia Appendix 5 for more details on speech recognition performance). Two studies [52, 61] described a qualitative assessment of VCAs' accuracy. One study [52] observed that the standard Google Search to perform better than voice-activated internet search performed with Google Assistant and Siri. The other study [61] reported on accuracy in assisting with social activities and showed all commercial VCAs to perform well at basic greeting activities, that Apple Siri and Amazon Alexa performed the best at email management but Apple Siri performed the worst at supporting social games. Moreover, Google Assistant performed the best at social game activities but the worst at social media management.

Technology acceptance

Seven of the 12 studies [50, 51, 53, 56-58, 60] reported technology acceptance findings, while the others (N=5) did not [52, 54, 55, 59, 61]. Three studies [51, 57, 58] reported technology acceptance through the System Usability Survey (SUS). One [58] reported a relatively high usability score (SUS score of M = 88/100), while one study [51] described better usability of its VCA for HPV in comparison to industry standards (i.e. SUS score of M=72/100). The latter also compared SUS scores between groups and found a higher score for participants who did not get the HPV vaccine (M=80/100), compared to those who did get the vaccine (M=77/100) and the control group (M=74/100). Also, the study found the score of Speech User Interface Service Quality (SUISQ) to be medium (M=4.29/7). The third study [57] reported a broader set of results and found, in addition to very good usability (SUS score: 82/100), a very good naturalness, information delivery, interpretability, and technology acceptance (all ≥8.25/10) of its VCA. Two studies [50, 60] reported different types of evaluation of technology acceptance. Thus, one study [50] reported good ease of use (5.4/7), acceptable expected capabilities (4.5/7), but low efficiency (3.3/7) of its VCA, while the other [60] described a positive user experience of its VCA with all User Experience Questionnaire (UEQ) constructs greater than 1.8/3. Finally, two studies reported a qualitative evaluation of their VCA, one [53] stating theirs to be "more accepted than rejected" in terms of user satisfaction, without giving more details, and the other [56] mentioning a generally positive assessment but a slowness in the processing of their VCA.

Methodology of the included studies

We included all types of measures that were present in more than one study, i.e., system accuracy measures, technology acceptance measures, behavioral measures, measures of attitude towards the target health behavior, and reported previous experience with technology.

The majority of the studies (N=10) did not report any behavioral measure [50-54, 56-58, 60, 61], while the two articles [55, 59] did. One [59] described the frequency of verbal responses not relevant to the system (i.e., non-matrix-vocabulary words), while the other [55] provided engagement and user performance (task completion, time to respond, points of difficulty, points of dropout, and quality of responses).

Half of the studies (N=6) did not report on any system measure [50, 51, 53, 56-58], while the other half reported either speech recognition performance measures (N=4) [54, 55, 59, 60] or criteria-based evaluation of the goodness of the VCA's repose (N=2)[52, 61]. In particular, four studies [54, 55, 59, 60] measured speech recognition performance, compared to human recognition. One of those [59] measured the accuracy of a diagnostic test score (i.e. speech reception threshold) compared to manually transcribed results. One [55] measured speech recognition percentage inferred from transcriptions of the interaction. One [54] compared the VCA to nurse practitioners' interpretation of patient's responses. Finally, one [60] gave more detailed results, reporting a confusion matrix, and speech recognition accuracy, precision, sensitivity, specificity, and f-measure as well as performance in task completion rate and prevention from security breaches.

Most of the 12 studies (N=7) [50, 51, 53, 56-58, 60] reported technology acceptance measures, while the remaining studies [52, 54, 55, 59, 61] did not. While two studies [51, 60] used a validated questionnaires only and two [53, 58] used an adapted questionnaires only, one study used both a validated questionnaire and an adapted questionnaire [57]. One paper [50] used an adapted questionnaire as well as qualitative feedback as acceptance measures. One study [56] reported qualitative feedback only.

The majority of the included studies (N=10) did not provide measures of attitude towards the target health behavior [52-61]. The two remaining papers [50, 51] provided validated questionnaires, and both focused on the attitude toward HPV vaccines. One paper [50] used the Parent Attitudes about Childhood Vaccines (PACV), and one [51] used the Carolina HPV Immunization Attitude and Belief Scale (CHIAS).

The majority of the included studies (N=8) also did not report controlling for participant's previous experience with technology [50-54, 57, 60, 61]. Of the remaining four studies, one study [59] reported that all study participants had no experience with smart speakers; one [58] informed that all study participants were familiar with mobile health applications; one [56] controlled for participants' smartphone ownership, use competence on Androids, iPhones, tablets, laptops, and desktop computers. Lastly, one paper [55] assessed the previous exposure of study participants to voice-based assistants was assessed but did not report on the assessments' results.

In general, risk bias varied importantly, from a minimum of 1 [61] to a maximum of 11.25 [51] (see more details in Multimedia Appendix 4).

Health Characteristics

Of the included studies, cancer was the most common health condition; two papers [50, 51] addressed cancer associated with Human Papillomavirus, whereas one study [52] addressed cancer associated with smoking. The next most common addressed conditions were diabetes (N=2) [53, 60] and heart failure (N=2) [54, 58]. Other discussed conditions were hearing-impairment [59], asthma [57], and Parkinson's disease [56]. Three papers addressed psychological conditions [55, 56, 61]. Specifically, they focused on dementia and autism [56], “intellectual disability” [55], and depression [61].

Three of the included studies [53, 58, 61] targeted elderly people, two targeted either parents of adolescents [50] or pediatric patients [51]. Other target populations were hearing-impaired individuals [59], smokers [52], asthma patients [57], glaucoma and diabetic patients [60], people with “intellectual disability” [55], and chronic heart failure patients [54]. One study [56] did not specify a particular target population.

The actual study participants consisted of the following populace: Healthy adults with at least one child under the age of 18 (N=16) [50]; healthy young adults between 18 and 16 years old (N=24) [51]; the authors themselves (N=2 [52]);; elderly people (N=10) [53]; chronic heart failure patients (N=14) [54]; adults with lifelong intellectual disability (N=9) [55]; adults recruited on campus (N=33) [56]; clinicians and researchers (N=16) [57]; healthy adults working regularly with senior patients (N=11) [58]; normal-hearing people (N=6) [59]; and adults affiliated to the university (N=33) [60]. One study [61] did not specified neither the type nor the number of participants.

Characteristics of Voice-based Conversational Agents

Eight studies [51, 53, 54, 56-60] named their VCA, while two studies [50, 55] did not specify any name. The two studies [52, 61] did not provide a name as they were evaluating existing commercially available VCA (i.e. Amazon Alexa, Cortana, Google Assistant, Apple Siri).
The majority of the included studies (N=7) did not report a description of the user interface of their VCAs [51-53, 55, 59, 61]. The remaining five papers did provide such a description [50, 56-58, 60].
The underlying architecture of the investigated VCAs was described in seven of the included studies [53, 54, 57-61], whereas three articles did not provide this information [52, 55, 56]. Two studies [50, 51] could not provide any architectural information given the nature of their study design (i.e., WOz).
When considering the devices used to test the VCA, we find that smartphones were the most utilized devices for data collection in the included studies (N=5) [52, 56-58, 60], followed by smart speakers (N=3)[53, 55, 59] and tablets (N=2) [50, 51]. Two studies [54, 61] did not explicitly specify which device they used for data collection.
The vast majority of the VCAs (N=10) were not commercially available [50, 51, 53-60] at the time of generation of this systematic literature review. In particular, one [56], reported the VCA to be available on Google Playstore at the time of publication but could not be found by the authors of this literature review at the time of reporting. Given that the other two studies tested on consumer VCA, we classified these papers as testing commercially available VCAs [52, 61].

Characteristics of Voice-Based Interventions

The interventions could be categorized in monitoring and/or support. Monitoring interventions refer to those focusing on health tracking (e.g., symptoms, medication adherence), while support interventions include targeted or on-demand information or alerts. This categorization was based on the classification of digital health interventions from the World Health Organization [62]. Five VCAs [50-52, 55, 61] exclusively focused on support, and three studies [54, 56, 59] exclusively focused on monitoring. Four studies investigated a VCA providing both monitoring and support [53, 57, 58, 60]. Monitoring activities were mainly implemented as active data capture and documentation (N=5) [53, 54, 57-60], whereas one study [57] also focused on self-monitoring of health or diagnostic data. One study [56] investigated self-monitoring of health or diagnostic data as the main monitoring activity.
Support services consisted mainly in delivering exclusively targeted health information based on health status (N=4) [50, 51, 55, 58, 60], whereas one study [58] also provided a look-up of health information. Three studies provided such a look-up of health information only [52, 53, 57], whereas two [53, 57] provided targeted alerts and reminders too. Lastly, one study delivered a support intervention in the form of task completion assistance [61] (see Multimedia Appendix 3 for more details on the interventions).

Discussion

Principal Findings

The goal of this systematic review was to summarize the available research on VCA for the prevention and management of chronic and mental conditions. Our investigation included 12 articles reporting mainly studies on the development and evaluation of a VCA, either in terms of system accuracy (e.g., speech recognition performance, appropriateness of VCA responses) or in terms of technology acceptance. Only one study reported on both aspects.
System accuracy referred to the ability of the VCA to interact with the participants, either in terms of

recognition performance or in terms of the ability to respond adequately to user queries. Speech recognition in VCA prototypes was mostly good or very good. The only relevant flaw revealed was a slowness in the VCA responses, which was reported in two of the selected studies [50, 56]. Commercial VCAs, although, not outperforming Google Search when the intervention involved look-up of health information, seem to have a specialization in supporting certain social activities (e.g. Apple Siri and Amazon Alexa for social media and office-related activities, Google Assistant for social games). These results suggest that there is a great potential for non-commercial VCAs, as they perform well in the domain for which they were built, while commercial VCAs are rather superficial in their health-related support.

Technology acceptance referred to all measures of the user's perception of the system (e.g., user experience, ease of use, efficiency of interaction). Despite the heterogeneity of technology acceptance measures, results showed good to very good performance. This suggests that dedicated VCAs are successfully capable of satisfying users' expectations when supporting prevention and/or management of chronic or mental conditions.

The majority of the included studies were published relatively recently, around 2019, and were fairly distributed between journal and conference or congress papers. Moreover, all studies were non-experimental and there was a general heterogeneity in the evaluation methods, especially in the user perception of the technology (i.e., user experience). Also, there was a general discrepancy between the target population and the actual sample recruited. In particular, although the VCAs studied were dedicated to the management or prevention of chronic and mental conditions, the evaluation was mainly conducted with healthy or convenience samples.

Considering the aspects mentioned above and the limited number of studies identified, it seems that the research in VCAs for chronic diseases and mental health is still in its infancy. Nevertheless, the results of almost all studies reporting system accuracy and technology acceptance are encouraging, especially for the developed VCAs, which inspires further development of this technology for the prevention and management of chronic and mental conditions

Related work

To the best of our knowledge, this is the only systematic literature review addressed VCAs specifically dedicated to the prevention and management of chronic diseases and mental illnesses. Only one scoping review appraised existing evidence on voice assistants for health and focused on interventions of healthy lifestyle behaviors in general [6]. Our findings are coherent with the review from Sezgin and colleagues [6] in a series of aspects. First, we also show that research in VCAs is still emerging, with studies including small samples and focusing on the feasibility of dedicating VCA for a specific health domain. Second, we also find a heterogeneous set of target populations and target health domain. However, our findings are in contrast with Sezgin et al [6] in the following aspects. First, we report studies mainly focusing on developing and evaluating the system in terms of system accuracy or technology acceptance: Sezgin et al [6] also describe efficacy tests but did not report on system accuracy. Third, the studies included in this review presented only VCA applications, while Sezgin et al [6] also included automated interventions via telephone. Finally, despite the preliminary character of the research, we include a risk bias assessment to formalize the importance of rigorous future research on VCAs for health.

In general, as we tried to include results explaining the technology acceptance of VCAs as a digital health intervention for prevention and management of chronic and mental conditions, our findings are more appropriate when concluding the state-of-the-art in evidence-based VCAs in this specific domain, rather than in healthy lifestyle behaviors in general.

Limitations

There are several limitations in our study, which may limit the generalization of our results. First, our search strategy focused on rather non-specific constructs (e.g., health), which may have led to the initial inclusion of a high number of unrelated literature, in addition to that concerning the main topic of this review (i.e., VCAs for chronic diseases and mental health). Given the infancy of this field, however, we chose a more inclusive strategy to avoid missing relevant literature for the analysis. Second, to evaluate a possible experimental bias of the studies, we followed the reporting guidelines suggested by JMIR and we chose the CONSORT-EHEALTH checklist. The risk bias varied importantly between the selected studies. This evaluation scheme may be regarded as unsuitable to evaluate the presented literature since none of the articles reports an experimental trial. An evaluation scheme capable of taking into account the pioneering character of the articles concerning the use of this technology for health-related applications could have enabled a more differentiated assessment.

Future work

The wide adoption of voice assistants worldwide and the interest in using them for healthcare purposes [30] generates great potential for effective implementation of scalable digital health interventions. There is, however, a lack of a clear implementation framework for VCAs. For instance, text-based and embodied conversational agents can currently be implemented using existing frameworks dedicated to digital health interventions [63-66] but there is, to the best of our knowledge, no such framework for VCAs. A platform for the development of VCAs dedicated to specific chronic or mental health conditions could encourage standardized implementations, which would be more comparable in their development and evaluation processes. Currently, it is possible to develop applications for consumer voice assistants (e.g., “skills” for Amazon Alexa or “actions” for Google Assistant). These products may, however, be of privacy [67] or safety [68] concerns. The academic community should therefore strive for the creation of such a platform, to foster the development of VCA for health.

The identified research provides diverse and general evaluation measures around the technology acceptance (or user experience in general) and no evaluation based on theoretical models of health behavior (e.g., intention of use). Thus, although the developed VCA might have been well received by the studied population samples, there is a need for a more systematic and comparable evaluation of the evidence-systems to understand which aspects of VCAs are best for user satisfaction. Future research should favor the use of multiple standardized questionnaires dedicated to voice user interfaces [69] for further exploration of the factors potentially influencing their effectiveness (e.g., rapport [70], intention of use [71]).

Moreover, only four papers [54, 55, 59, 60] reported comparing the accuracy of the VCA’s interpretation of participants’ responses to humans’. Although it was limited to speech recognition, they were the only cases of human-machine comparison. To verify the suitability of VCAs as an effective and scalable alternative to healthcare practitioners, more research should compare not only the system accuracy but also the general performance of this type of digital health intervention in comparison to standard in-person healthcare.

Finally, all papers conducted laboratory experiments and were focusing on short-term performance and/or technology acceptance. Even if this evidence shows the feasibility of VCAs for healthcare, it does not provide evidence on the actual effectiveness of VCAs in assisting patients in managing their chronic and mental health conditions compared to standard practices. Future research should provide evidence on complementary short-term and long-term measurements of technology acceptance, and behavioral and health outcomes associated with the use of VCAs.

Conclusions

This study provides a systematic review of VCAs for the prevention and management of chronic and mental conditions. Out of 7,170 prescreened articles, we included and analyzed12 articles reporting studies either on the development and evaluation of a VCA, or on the criteria-based evaluation of commercial VCAs. We found that all studies were non-experimental and there was a general heterogeneity in the evaluation methods. Considering the recent publication date of the included articles, we conclude that this field is still in its infancy. The results of almost all studies on the performance of the system and the experiences of users are, however, encouraging. Even if the evidence provided in this systematic review shows the feasibility of VCAs for healthcare, current research does not provide any insight on the actual effectiveness of VCAs in assisting patients in managing their chronic and mental conditions. Future research should therefore especially focus on the investigation of health and behavioral outcomes, together with relevant technology acceptance outcomes associated with the use of VCAs.

Acknowledgments

This study is supported by the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise program.

Authors' Contributions

CB, EF, and TK were responsible for the study design and search strategy. CB and RK were responsible for the screening and data extraction. CB, RK, and TS were responsible for the data analysis. CB, RK, TS, and FB were responsible for the first draft. All authors were responsible for critical feedback and final revisions of the manuscript. TS and RK share second authorship. FB and TK share last authorship.

Conflicts of Interest

All authors are affiliated with the Center for Digital Health Interventions (CDHI) (www.c4dhi.org), a joint initiative of the Department of Management, Technology, and Economics at ETH Zurich and the Institute of Technology Management at the University of St. Gallen, which is funded in part by the Swiss health insurer CSS. EF and TK are also co-founders of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. Neither CSS nor Pathmate Technologies were involved in any way in this manuscript.

Abbreviations

CHIAS: Carolina HPV Immunization Attitude and Belief Scale
HPV: human papillomavirus
JMIR: Journal of Medical Internet Research
PACV: Parent Attitudes about Childhood Vaccines
RCT: randomized controlled trial
SQUISQ: Speech User Interface Service Quality
UEQ: User Experience Questionnaire
VCA: voice-based conversational agent
WOz: Wizard-of-Oz

Multimedia Appendix 1

Study protocol.

Multimedia Appendix 2

Search terms per construct (syntax used in PubMed Medline).

Multimedia Appendix 3

Complete list of characteristics of the included studies.

Multimedia Appendix 4

Risk-of-bias assessment.

Multimedia Appendix 5

Main characteristics of the included studies.

References

1. WHO. World Health Statistics 2020: Monitoring Health for the SDGs, Sustainable Development Goals. 2020 2020. Report No.
2. WHO. Suicide in the World: Global Health Estimates. 2019 2019.
3. Kvedar JC, Fogel AL, Elenko E, Zohar D. Digital medicine's march on chronic disease. Nat Biotechnol. 2016 2016;34(3):239-46. doi: 10.1038/nbt.3495.
4. 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 2015;17(2):e52. doi: 10.2196/jmir.3951.
5. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res. 2018 2018;107:73-8. doi: 10.1016/j.jpsychires.2018.10.006.
6. Sezgin E, Militello LK, Huang Y, Lin S. A scoping review of patient-facing, behavioral health interventions with voice assistant technology targeting self-management and healthy lifestyle behaviors. Transl Behav Med. 2020 2020;10(3):606-28. doi: 10.1093/tbm/ibz141.
7. 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 2018;25(9):1248-58. doi: 10.1093/jamia/ocy072.
8. Schachner T, Keller R, von Wangenheim F. Artificial intelligence based conversational agents for chronic conditions: A systematic literature review (preprint). J Med Internet Res. 2020 2020. doi: 10.2196/20701.
9. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y-L, et al. Conversational agents in health care: Scoping review and conceptual analysis. J Med Internet Res. 2020 2020;22(8):e17158. doi: 10.2196/17158.
10. Ammari T, Kaye J, Tsai JY, Bentley F. Music, search, and IoT: How people (really) use voice assistants. ACM Trans Comput Hum Interact. 2019 2019;26(3):1-28. doi: 10.1145/3311956.
11. Hoy MB. Alexa, Siri, Cortana, and more: an introduction to voice assistants. Medical reference services quarterly. 2018;37(1):81-8.
12. Liu B, Sundar SS. Should machines express sympathy and empathy? Experiments with a health advice chatbot. Cyberpsychol Behav Soc Netw. 2018 2018;21(10):625-36. doi: 10.1089/cyber.2018.0110.
13. Schwartzman CM, Boswell JF. A narrative review of alliance formation and outcome in text-based telepsychotherapy. Pract Innov (Wash, DC). 2020 2020;5(2):128-42. doi: 10.1037/pri0000120.
14. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns. 2005 2005;59(1):21-30. doi: 10.1016/j.pec.2004.09.008.
15. Horvath AO, Luborsky L. The role of the therapeutic alliance in psychotherapy. J Consult Clin Psychol. 1993 1993;61(4):561-73. doi: 10.1037/0022-006x.61.4.561.

16. Leach MJ. Rapport: a key to treatment success. Complement Ther Clin Pract. 2005 2005;11(4):262-5. doi: 10.1016/j.ctcp.2005.05.005.
17. Mead N, Bower P. Patient-centred consultations and outcomes in primary care: a review of the literature. Patient Educ Couns. 2002 2002;48(1):51-61. doi: 10.1016/s0738-3991(02)00099-x.
18. Martin DJ, Garske JP, Davis MK. Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. J Consult Clin Psychol. 2000 2000;68(3):438-50. doi: 10.1037/0022-006x.68.3.438.
19. Nass C, Steuer J, Tauber ER, editors. Computers are social actors. Proceedings of the SIGCHI conference on Human factors in computing systems; 1994.
20. Fogg BJ. Persuasive computers: Perspectives and research directions. Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '98; 1998. New York, New York, USA: ACM Press; 1998.
21. Cho E. Hey Google, can I ask you something in private? Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19; 2019. New York, New York, USA: ACM Press; 2019.
22. Kim K, Norouzi N, Losekamp T, Bruder G, Anderson M, Welch G. Effects of patient care assistant embodiment and computer mediation on user experience. 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR); 2019: IEEE; 2019.
23. Barata M, Galih Salman A, Faahakhododo I, Kanigoro B. Android based voice assistant for blind people. Libr hi tech news. 2018 2018;35(6):9-11. doi: 10.1108/lhtn-11-2017-0083.
24. Balasuriya SS, Sitbon L, Bayor AA, Hoogstrate M, Brereton M. Use of voice activated interfaces by people with intellectual disability. Proceedings of the 30th Australian Conference on Computer-Human Interaction; 2018. New York, NY, USA: ACM; 2018.
25. Masina F, Orso V, Pluchino P, Dainese G, Volpato S, Nelini C, et al. Voice assistants are for (almost) everybody: investigating the accessibility of speech-based interactive systems with impaired users (Preprint). J Med Internet Res. 2020 2020. doi: 10.2196/18431.
26. Sezgin E, Huang Y, Ramtekkar U, Lin S. Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic. NPJ Digit Med. 2020 2020;3:122. doi: 10.1038/s41746-020-00332-0.
27. Pulido MLB, Hernández JBA, Ballester MÁF, González CMT, Mekyska J, Smékal Z. Alzheimer's disease and automatic speech analysis: A review. Expert Syst Appl. 2020 2020;150(113213):113213. doi: 10.1016/j.eswa.2020.113213.
28. Wang J, Zhang L, Liu T, Pan W, Hu B, Zhu T. Acoustic differences between healthy and depressed people: a cross-situation study. BMC Psychiatry. 2019 2019;19(1):300. doi: 10.1186/s12888-019-2300-7.
29. Mota NB, Copelli M, Ribeiro S. Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance. NPJ Schizophr. 2017 2017;3:18. doi: 10.1038/s41537-017-0019-3.
30. Kinsella B, Mutchler A. Voice Assistant Consumer Adoptoin in Healthcare. 2019 2019. Report No.
31. PRC. Nearly half of Americans use digital voice assistants, mostly on their smartphones. 2017 [updated 2017/12/12]; Available from: https://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/.
32. Chung AE, Griffin AC, Selezneva D, Gotz D. Health and fitness apps for hands-free voice-activated assistants: Content analysis. JMIR MHealth UHealth. 2018 2018;6(9):e174. doi: 10.2196/mhealth.9705.
33. Aiva. Aiva: Virtual Health Assistant. 2020 [11/23/2020]; Available from:

https://www.aivahealth.com/.
34. Orbita. Orbita AI – Leader in Conversational AI for Healthcare. [cited 11/23/2020]; Available from: https://orbita.ai/.
35. Omron. OMRON Health skill for Amazon Alexa. 2020 [updated 2019/6/1823/11/2020]; Available from: https://omronhealthcare.com/alexa/.
36. Wellpepper. Sugarpod.io. 2013-2017 [11/23/2020]; Available from: http://sugarpod.io/.
37. MFMER. Skills from mayo clinic. 2020 [updated 2020/5/611/23/2020]; Available from: https://www.mayoclinic.org/voice/apps.
38. López G, Quesada L, Guerrero LA. Alexa vs. Siri vs. Cortana vs. Google assistant: A comparison of speech-based natural user interfaces. Advances in Intelligent Systems and Computing. Cham: Springer International Publishing; 2018. p. 241-50. doi: 10.1007/978-3-319-60366-7_23
39. Pradhan A, Lazar A, Findlater L. Use of intelligent voice assistants by older adults with low technology use. ACM Trans Comput Hum Interact. 2020 2020(3373759). doi: 10.1145/3373759.
40. HDRO. Human Development Data (1990-2018). 1990-2018 [11/23/2020]; Available from: http://hdr.undp.org/en/data.
41. Rosenberg S. Smartphone ownership is growing rapidly around the world, but not always equally. 2019 2019/2/5. Available from: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/.
42. Holden RJ, Karsh B-T. The technology acceptance model: its past and its future in health care. J Biomed Inform. 2010 2010;43(1):159-72. doi: 10.1016/j.jbi.2009.07.002.
43. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 2015;349(jan02 1):g7647-g. doi: 10.1136/bmj.g7647.
44. WHO. Fact sheets: Chronic Diseases. 2014. Available from: https://www.who.int/topics/chronic_diseases/factsheets/en/.
45. WHO. Fact sheets: Mental disorders. 2019. Available from: https://www.who.int/news-room/fact-sheets/detail/mental-disorders.
46. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960 1960;20(1). doi: 10.1177/001316446002000104
47. Altman DG. Practical statistics for medical research. London, England: Chapman and Hall; 1999 1999. ISBN: 9780412819308.
48. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res. 2014;16(2):e40. doi: 10.2196/jmir.2952.
49. Schulz KF, Altman DG, Consort MD. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010 2010;2010;340:c332. doi: 10.1136/bmj.c332.
50. Amith M, Zhu A, Cunningham R, Lin R, Savas L, Shay L, et al. Early usability assessment of a conversational agent for HPV vaccination. Stud Health Technol Inform. 2019 2019;257:17-23. PMID: 30741166.
51. Amith M, Lin R, Cunningham R, Wu QL, Savas LS, Gong Y, et al. Examining potential usability and health beliefs among young adults using a conversational agent for HPV vaccine counseling. AMIA Summits Transl Sci Proc. 2020 2020;2020:43-52. PMID: 32477622.
52. Boyd M, Wilson N. Just ask Siri? A pilot study comparing smartphone digital assistants and laptop Google searches for smoking cessation advice. PLoS One. 2018

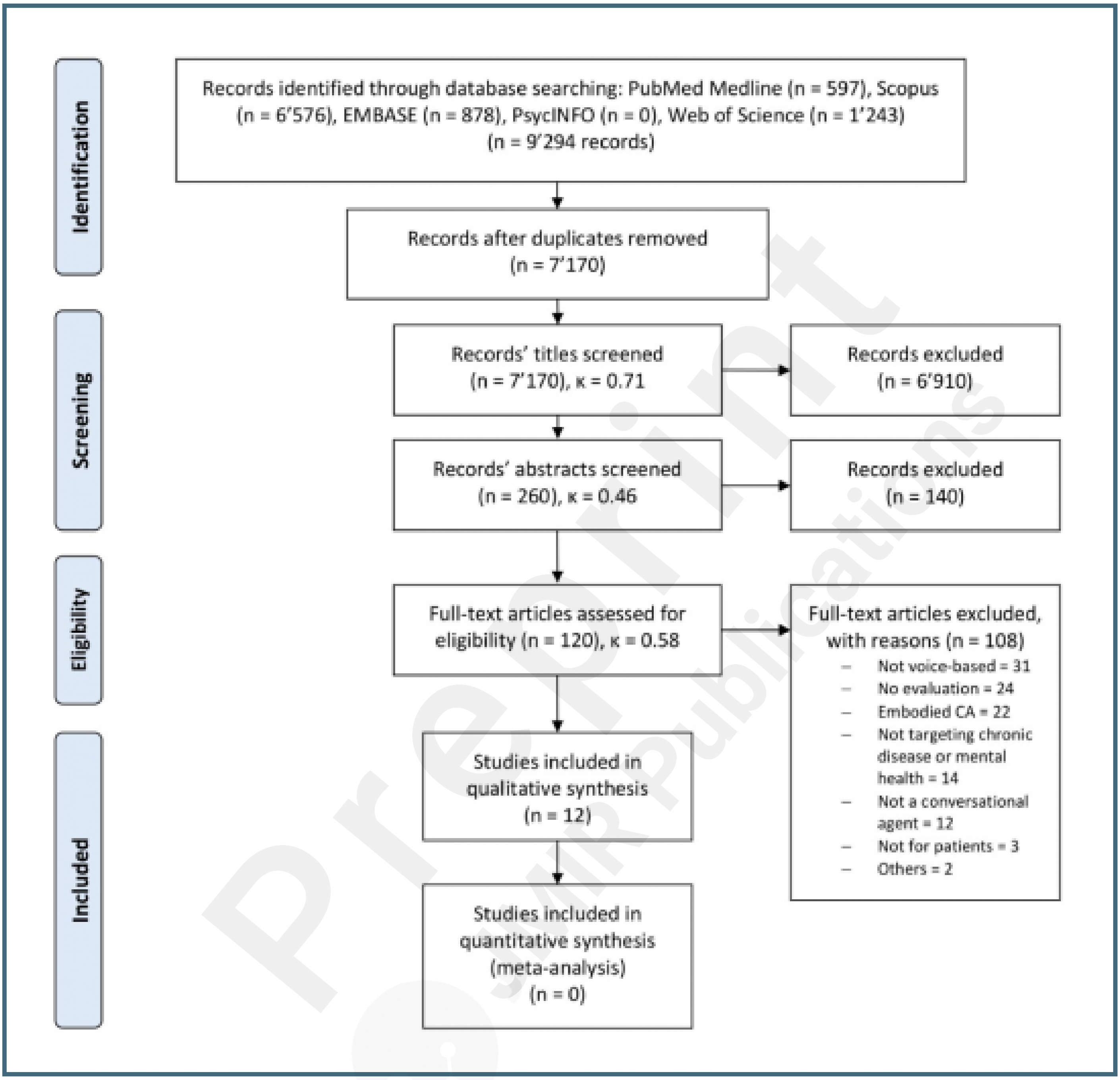
2018;13(3):e0194811. doi: 10.1371/journal.pone.0194811.
53. Cheng A, Raghavaraju V, Kanugo J, Handrianto YP, Shang Y. Development and evaluation of a healthy coping voice interface application using the Google home for elderly patients with type 2 diabetes. 2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC); 2018: IEEE; 2018. doi: 10.1109/CCNC.2018.8319283
54. Galescu L, Allen J, Ferguson G, Quinn J, Swift M. Speech recognition in a dialog system for patient health monitoring. 2009 IEEE International Conference on Bioinformatics and Biomedicine Workshop; 2009: IEEE; 2009. doi: 10.1109/BIBMW.2009.5332111.
55. Greuter S, Balandin S, Watson J. Social games are fun: Exploring social interactions on smart speaker platforms for people with disabilities. Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts - CHI PLAY '19 Extended Abstracts; 2019. New York, New York, USA: ACM Press; 2019. doi: 10.1145/3341215.3356308.
56. Ireland D, Atay C, Liddle J, Bradford D, Lee H, Rushin O, et al. Hello Harlie: Enabling speech monitoring through chat-bot conversations. Stud Health Technol Inform. 2016 2016;227:55-60. PMID: 27440289.
57. Kadariya D, Venkataramanan R, Yip HY, Kalra M, Thirunarayanan K, Sheth A. KBot: Knowledge-enabled personalized chatbot for asthma self-management. Proc Int Conf Smart Comput SMARTCOMP. 2019 2019;2019:138-43. doi: 10.1109/smartcomp.2019.00043.
58. Lobo J, Ferreira L, Ferreira AJS. CARMIE: A conversational medication assistant for heart failure. Int j e-health med commun. 2017 2017;8(4):21-37. doi: 10.4018/ijehmc.2017100102.
59. Ooster J, Moreta PNP, Bach J-H, Holube I, Meyer BT. “computer, test my hearing”: Accurate speech audiometry with smart speakers. Interspeech 2019; 2019. ISCA: ISCA; 2019.
60. Rehman UU, Chang DJ, Jung Y, Akhtar U, Razzaq MA, Lee S. Medical Instructed Real-time Assistant for patient with glaucoma and diabetic conditions. Appl Sci (Basel). 2020 2020;10(7):2216. doi: 10.3390/app10072216.
61. Reis A, Paulino D, Paredes H, Barroso I, Monteiro MJ, Rodrigues V, et al. Using intelligent personal assistants to assist the elderlies An evaluation of Amazon Alexa, Google Assistant, Microsoft Cortana, and Apple Siri. 2018 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW); 2018: IEEE; 2018.
62. WHO. Classification of digital health interventions v1.0: A shared language to describe the uses of digital technology for health. 2018 WHO/RHR/19.06.
63. CDHI. MobileCoach. 2020; Available from: https://www.mobile-coach.eu/.
64. Filler A, Kowatsch T, Haug S, Wahle F, Staake T, Fleisch E. MobileCoach: A novel open source platform for the design of evidence-based, scalable and low-cost behavioral health interventions: Overview and preliminary evaluation in the public health context. 2015 Wireless Telecommunications Symposium (WTS); 2015: IEEE; 2015.
65. RAG. Relational Agents Group. 2019 [11/23/2020]; Available from: http://relationalagents.com.
66. Bickmore TW, Schulman D, Sidner CL. A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. J Biomed Inform. 2011 2011;44(2):183-97. doi: 10.1016/j.jbi.2010.12.006.
67. 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 2018;20(9):e11510. doi: 10.2196/11510.
68. Chung H, Iorga M, Voas J, Lee S. Alexa, can I trust you? Computer (Long Beach Calif). 2017 2017;50(9):100-4. doi: 10.1109/MC.2017.3571053.
69. Kocabalil AB, Laranjo L, Coiera E. Measuring user experience in conversational

interfaces: A comparison of six questionnaires. HCI 2018; 2018: BCS Learning & Development; 2018. doi: 10.14236/ewic/HCI2018.21
70. Falkenström F, Hatcher RL, Skjulsvik T, Larsson MH, Holmqvist R. Development and validation of a 6-item working alliance questionnaire for repeated administrations during psychotherapy. Psychol Assess. 2015 2015;27(1):169-83. doi: 10.1037/pas0000038.
71. Venkatesh, Thong, Xu. Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. MIS Q. 2012 2012;36(1):157. doi: 10.2307/41410412.

Supplementary Files

Figures

PRISMA flow diagram of included studies.



Multimedia Appendixes

Study protocol.
URL: https://asset.jmir.pub/assets/5db5da3f2730820fff995234a3782eb9.pdf

Search terms per construct (syntax used in PubMed Medline).
URL: https://asset.jmir.pub/assets/251aa27a24e338338316d4d4923744f3.pdf

Complete list of characteristics of the included studies.
URL: https://asset.jmir.pub/assets/fb70b774fdc75bab1e1f8716d61cd502.pdf

Risk-of-bias assessment.
URL: https://asset.jmir.pub/assets/15e12ea66a66d5cafdc08c906301cceb.pdf

Main characteristics of the included studies.
URL: https://asset.jmir.pub/assets/408f9be91e7d748e0fdc95c255defe94.pdf

TOC/Feature image for homepages

Untitled.