Protocol

Conversational Agents for Health and Well-being Across the Life Course: Protocol for an Evidence Map

Mara Pereira Guerreiro[1,2], PhD; Leonardo Angelini[3], PhD; Helga Rafael Henriques[1], PhD; Mira El Kamali[3], MSc; Cristina Baixinho[1,4], PhD; João Balsa[5], PhD; Isa Brito Félix[1], MSc; Omar Abou Khaled[3], PhD; Maria Beatriz Carmo[5], PhD; Ana Paula Cláudio[5], PhD; Maurizio Caon[3], PhD; Karl Daher[3], MSc; Bruno Alexandre[6], BSc; Mafalda Padinha[7], BSc; Elena Mugellini[3], PhD

[1]Nursing Research, Innovation and Development Centre of Lisbon, Nursing School of Lisbon, Lisbon, Portugal

[2]Centro de Investigação Interdisciplinar Egas Moniz, Instituto Universitário Egas Moniz, Monte de Caparica, Portugal

[3]University of Applied Sciences and Arts Western Switzerland, Fribourg, Switzerland

[4]CiTechare, Leiria, Portugal

[5]LASIGE, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

[6]Nursing School of Lisbon, Lisbon, Portugal

[7]Instituto Universitário Egas Moniz, Monte de Caparica, Portugal

Corresponding Author:
Mara Pereira Guerreiro, PhD
Nursing Research, Innovation and Development Centre of Lisbon
Nursing School of Lisbon
Avenida Professor Egas Moniz
Lisbon
Portugal
Phone: 351 217913400 ext 23507
Email: mara.guerreiro@esel.pt

Abstract

Background: Conversational agents, which we defined as computer programs that are designed to simulate two-way human conversation by using language and are potentially supplemented with nonlanguage modalities, offer promising avenues for health interventions for different populations across the life course. There is a lack of open-access and user-friendly resources for identifying research trends and gaps and pinpointing expertise across international centers.

Objective: Our aim is to provide an overview of all relevant evidence on conversational agents for health and well-being across the life course. Specifically, our objectives are to identify, categorize, and synthesize—through visual formats and a searchable database—primary studies and reviews in this research field.

Methods: An evidence map was selected as the type of literature review to be conducted, as it optimally corresponded to our aim. We systematically searched 8 databases (MEDLINE; CINAHL; Web of Science; Scopus; the Cochrane, ACM, IEEE, and Joanna Briggs Institute databases; and Google Scholar). We will perform backward citation searching on all included studies. The first stage of a double-stage screening procedure, which was based on abstracts and titles only, was conducted by using predetermined eligibility criteria for primary studies and reviews. An operational screening procedure was developed for streamlined and consistent screening across the team. Double data extraction will be performed with previously piloted data collection forms. We will appraise systematic reviews by using A Measurement Tool to Assess Systematic Reviews (AMSTAR) 2. Primary studies and reviews will be assessed separately in the analysis. Data will be synthesized through descriptive statistics, bivariate statistics, and subgroup analysis (if appropriate) and through high-level maps such as scatter and bubble charts. The development of the searchable database will be informed by the research questions and data extraction forms.

Results: As of April 2021, the literature search in the eight databases was concluded, yielding a total of 16,351 records. The first stage of screening, which was based on abstracts and titles only, resulted in the selection of 1282 records of primary studies and 151 records of reviews. These will be subjected to second-stage screening. A glossary with operational definitions for supporting the study selection and data extraction stages was drafted. The anticipated completion date is October 2021.

Conclusions: Our wider definition of a conversational agent and the broad scope of our evidence map will explicate trends and gaps in this field of research. Additionally, our evidence map and searchable database of studies will help researchers to avoid fragmented research efforts and wasteful redundancies. Finally, as part of the Harnessing the Power of Conversational e-Coaches for Health and Well-being Through Swiss-Portuguese Collaboration project, our work will also inform the development of an international taxonomy on conversational agents for health and well-being, thereby contributing to terminology standardization and categorization.

International Registered Report Identifier (IRRID): DERR1-10.2196/26680

Introduction

In 2016, noncommunicable diseases (NCDs) accounted for 40.5 million deaths worldwide, which corresponded to 71% of deaths worldwide. The top 4 NCDs are cardiovascular diseases, cancers, diabetes, and chronic lung diseases [1]. NCDs can be prevented by adopting a healthy lifestyle. For example, a European multicohort study, which was conducted from 1991 to 2006 and included 116,043 people who were free of major NCDs at baseline, suggested that various healthy lifestyle profiles yield gains in life years without major NCDs, including type 2 diabetes, coronary heart disease, stroke, cancer, asthma, and chronic obstructive pulmonary disease [2].

Changing and sustaining health behaviors, which are integral to both the prevention and self-management of NCDs, are known to be challenging and resource intensive [3]. Digitalization and automation remove time and place restrictions, thereby broadening access to lifestyle and self-management interventions in a potentially cost-effective manner. For example, full economic evaluations of interventions that use the internet, mobile devices, or computers for the prevention and control of type 2 diabetes have demonstrated high cost-effectiveness, even though they were not fully automated [4].

The European Blueprint on Digital Transformation of Health and Care for the Ageing Society reflects a common vision that key stakeholders have on the role of innovation in changing health and care provision among older populations [5]. The priority topics encompassed by this policy vision are disease prevention, personalized health and care, and digital tools for citizen empowerment and person-centered care. A total of 12 blueprint personas have been created based on the health and care needs of people across the life course, ranging from children to persons aged ≥80 years [5].

Digital technology that mimics human communication is suitable for different age populations and populations with different literacy levels and is arguably more engaging for long-term use. A scan of this landscape has revealed a considerable body of scientific literature, although no agreements have emerged on the definition of so-called conversational agents. For instance, some authors consider conversational agents to be software capable of natural language processing [6]. However, others have used broader definitions that encompass agents that use predefined text options as inputs but exclude embodied agents that use nonverbal communication [7].

For the purpose of our review, we defined conversational agents as computer programs that are designed to simulate two-way human conversation by using language (speech or text) and are potentially supplemented with nonlanguage modalities. We believe that conversational agents for health and well-being should be further characterized according to the health intervention (eg, target population, design, the entity on which the intervention is carried out, and duration), the agent (eg, embodiment, role, and delivery channel), and the conversation (eg, input and output options, dialogue engine, and sentiment detection). We derived these ideas from literature [7-11], the international classification of health interventions [12], our interdisciplinary experience [13-16], and discussions within the research team.

This study is part of the Harnessing the Power of Conversational e-Coaches for Health and Well-being Through Swiss-Portuguese Collaboration (eCCo) project [17], which encompasses an evidence map and the subsequent development of an international taxonomy on conversational agents for health and well-being via a scientific consensus method that will be informed by our literature review. In addition to this purpose, the evidence map independently serves a much-needed research endeavor—fostering collaboration in the field through an open-access resource. In their review on conversational interfaces for health, Xing et al [18] highlighted the need to improve collaboration among stakeholders in research and patent activities. To our knowledge, there is no open, searchable database on conversational agents for health—a resource that could foster collaboration by pinpointing expertise across international centers and networks. Such collaboration can help with tackling research fragmentation and duplication.

Our aim is to provide an overview of all relevant evidence on conversational agents for health and well-being across the life course. Specifically, our objectives are to identify, categorize, and synthesize primary studies and reviews on this topic by focusing on the following research questions:

- What is the nature of literature on conversational agents for health and well-being (eg, information source, research group, and study characteristics)?

- What are the characteristics of health interventions based on conversational agents (eg, setting, target population, intervention target, duration, and frequency)?
- What are the characteristics of the automated conversations conducted in health interventions (eg, interaction input and output and dialogue engine)?
- What are the characteristics of the agents used in health interventions (eg, embodiment, emotions, role, and delivery channel)?

Methods

Evidence Map

An evidence map is "a systematic search of a broad field to identify gaps in knowledge and/or future research needs that presents results in a user-friendly format, often a visual figure or graph, or a searchable database" [19]. These reviews typically encompass different types of studies, such as reviews and primary studies. They rely on a systematic search strategy, conducting screening based on explicit eligibility criteria, and conducting data extraction in a structured format. Critical appraisal may be performed, but it is not required [19].

Searching

Keyword Selection and Initial Database Query

To comprehensively identify relevant keywords, we resorted to using a purposive sample of 13 literature reviews [6,7,9,10,18,20-29] and a review protocol [30] related to conversational agents for health and well-being. A total of 318 keywords were extracted. The removal of duplicates resulted in 220 keywords.

Keywords were categorized into tentative domains and tested in MEDLINE; we resorted to using PubMed as the interface. The search process was documented and iteratively optimized [31] to yield a compromise between feasibility and completeness. This led to the choice of using the following two final keyword domains: K1 (variations of conversational agent–related terms) and K2 (variations of health-related and well-being–related terms). We expanded the K1 domain by including all variations and combinations of the terms *agent* (ie, *bot*, *robot*, *assistant*, *coach*, *companion*, *system*, *avatar*, and *entity program*) and *conversational* (ie, *talking*, *voice*, *communication*, *social*, *dialogue*, and *utterance*). We also included terms in the K1 domain related to popular commercial conversational agents, such as *Google Home*, *Google Assistant*, *Cortana*, *Alexa*, and *Siri*.

The search strategy encompassed (when applicable) plural forms of keywords and variations at the end of keywords, which were indicated with the wildcard asterisk (ie, "*"). We accounted for variations in the middle of phrases and hyphenation by using similar vocabulary (eg, *talk bot* and *talkative bot*, *ecoach* and *e-coach*, etc). We ended up with 265 keywords for the K1 domain and 13 keywords for the K2 domain.

There was ambiguity between the terms *Amazon Alexa assistant* and *alexa fluor compounds*. Therefore, a third keyword domain was developed (K3), which consisted of variations of *alexa fluor compounds* to be excluded from the search query. This domain limited the number of irrelevant results through the use of database syntax.

The search string for the search conducted on MEDLINE via PubMed is depicted in Multimedia Appendix 1. The search was restricted to titles and abstracts only.

Database Selection

Initially, we listed the data sources used by the aforementioned sample of studies [6,7,9,10,18,20-30]. This led to a set of the following nine potentially useful scientific literature databases: MEDLINE; CINAHL; Web of Science; Scopus; the Cochrane, ACM, IEEE, and Joanna Briggs Institute (JBI) databases; and Google Scholar. These were then analyzed in terms of their coverage and suitability.

Both ACM and IEEE publish computer science conference proceedings, and these were relevant to our evidence map and supplemented our health data sources. Although Scopus and Web of Science cover most journal publications from ACM and IEEE, their indexing of conference proceedings is poorer [32]. Therefore, from a coverage standpoint, it would make sense to retain these databases.

In terms of the search quality of the data sources, the work of Gusenbauer and Haddaway [33] endorsed the choice of using MEDLINE, CINAHL, Web of Science, Scopus, Cochrane, and ACM. These authors did not consider Google Scholar to be an appropriate principal search system, since it does not allow for the use of Boolean queries and does not provide consistent results over time [33].

After taking both coverage and search qualities into account, we decided to retain Google Scholar and ACM in the final list of databases. When compared to Scopus and Web of Science, Google Scholar is still the most far-reaching source [34]. Scopus and Web of Science exhibit indexing lags and may miss the latest publications, unlike Google Scholar [35,36]. However, we discarded IEEE from the list on the grounds of its limited search capabilities, as we were unable to search this database for all the proposed keywords.

The decision to exclude grey literature was dictated by our available resources. A definition of grey literature was put forward during the 1997 International Conference on Grey Literature in Luxembourg and was expanded in 2004 (in New York) as "information produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers, i.e. where publishing is not the primary activity of the producing body" [37]. Recently, Garousi et al [38] proposed a wider definition for grey literature in the software engineering field and grouped grey literature into 3 tiers according to expertise (ie, the established knowledge of the content producer) and outlet control (ie, content production in conformance with explicit and transparent criteria). The tiers encompass content from blogs, tweets, and news articles; presentations, and government reports. Regardless of the definition, grey literature would add to the predictably extensive amount of formal literature and was deemed to be of uncertain value in light of the review's aim. Grey literature would also require additional resources for analysis.

Testing the Query on the Remaining Databases

The PubMed search query was used in CINAHL (via EBSCO [Elton B. Stephens Company]); Web of Science; Scopus; and the Cochrane (via EBSCO), ACM, and JBI databases; minor adjustments were made [31]. The restriction to a title and abstract search was maintained in these databases, with the exception of the JBI database, which does not allow for such searches.

The Google Scholar query, which is presented in Multimedia Appendix 1, comprised short forms of the K1 and K2 domain terms. This search was limited to literature published from 2020 onward and those without patents and citations. The search results were sorted by relevance. As recommended by Haddaway et al [39], we retrieved only the first 300 results.

Citation Searching

We will conduct a backward citation search by manually searching the reference lists of all articles included in the evidence map. Forward citation searching by using a citation index to identify studies that cite included articles [40] was deemed unfeasible in light of the project resources.

Selection of Studies

Rayyan (Rayyan Systems Inc)—a collaborative web application—was used to streamline the selection of studies [41]. It supports the screening and coding of studies, documents reviewers' decisions by using tags, and allows for the organization of records via filters.

Teams of 2 researchers will independently screen retrieved records by using predetermined eligibility criteria for primary studies and reviews, which are detailed in Textboxes 1 and 2. Meeting all of the inclusion criteria will be a requirement for an article to be selected for the evidence map. Selected articles must also not meet any of the exclusion criteria.

Textbox 1. Eligibility criteria for primary studies on conversational agents for health and well-being.

Inclusion criteria

- Primary studies that focus on persons of all ages regardless of their health status

- Presenting a computer program that is able to simulate two-way human conversation for a health-related purpose or general well-being–related purpose by using language (speech or text) and is potentially supplemented with nonlanguage modalities, regardless of the input and output options

- Reporting the design, development, evaluation, or implementation of conversational agents regardless of the involvement of human users and study design

Exclusion criteria

- Articles focused solely on caregivers, health care professionals, or the education of health care professionals or students

- Articles that do not concomitantly report information on the following three components: the health intervention, the agent, and the conversational capabilities (eg, articles focused on individual features only, such as speech recognition)

- Agents without automated conversational capabilities (eg, Wizard of Oz tool)

- Press articles

- Unavailable full text

- Articles written in languages other than English, Italian, French, Portuguese, and Spanish

- Commentaries, opinion papers, position papers, study protocols, or any article not presenting primary research (eg, discussing the intention to develop a conversational agent)

- Conference abstracts

Textbox 2. Eligibility criteria for reviews on conversational agents for health and well-being.

Inclusion criteria

- All review designs that focus on primary studies on human participants regardless of their health status

- Reviews comprised of studies that present a computer program that is able to simulate human conversation for a health-related purpose or general well-being–related purpose by using language and is potentially supplemented with nonlanguage modalities, regardless of the input and output options

- Reporting the design, development, evaluation, implementation, or funding of conversational agents regardless of the involvement of human users and study design

Exclusion criteria

- Reviews that include conversational agents for the education of health care professionals or students or another nonhealth purpose

- Reviews including studies on nonconversational agents or those without automated conversational capabilities (eg, Wizard of Oz tool)

- Review protocols

- Unavailable full text

- Articles written in languages other than English, Italian, French, Portuguese, and Spanish

The screening procedure was piloted at the commencement of this stage; we used a set of primary studies and reviews. In the first stage, we reviewed the titles and abstracts of retrieved records. Discrepancies in inclusion and exclusion decisions were resolved by a third reviewer. In the second stage we will focus on the full-text review of the records selected in the first stage. Discrepancies in inclusion and exclusion decisions at this stage will be resolved by discussion between the reviewer pairs, and if a consensus is not reached, a third researcher will be involved.

Data Extraction

Data collection forms for primary studies and reviews were designed using Microsoft Excel spreadsheets. In addition to general information such as article ID numbers (unique identifiers), titles, and aims, we will extract information related to our research questions. Moreover, we will extract definitions of conversational agents if they are provided. The forms will be piloted with a set of primary studies and reviews to ensure that they capture relevant information comprehensively. Multimedia Appendix 2 presents variables that were preliminarily included in the Excel spreadsheet.

Teams of 2 researchers will independently extract data from included records. Potential discrepancies will be resolved via consultation with a third researcher.

Critical Appraisal

As previously explained, critical appraisal is recommended for evidence maps but is not mandatory [19]. Therefore, based on project resources, we will conduct the critical appraisal of systematic reviews, but this will not be done for the anticipated large number of primary studies. A Measurement Tool to Assess Systematic Reviews (AMSTAR) 2—a revised 16-item version of AMSTAR—will be used to evaluate the quality of included systematic reviews [42]. AMSTAR 2 takes longer to apply than AMSTAR; however, both have higher levels of interrater

reliability compared to those of similar tools, such as the Risk of Bias Assessment Tool for Systematic Reviews (ROBIS) [43]. Compared to the ROBIS, AMSTAR 2 is easier to apply. Further, guidance on using AMSTAR 2 is clearer and simpler, which promotes its use by nonexperienced reviewers [44].

Data Synthesis

We will assess primary and secondary studies separately in the analysis. Each data set will be subjected to a descriptive analysis to summarize the characteristics of included studies. This will be guided by our research questions. The bivariate exploration of data will be conducted, as appropriate. Subgroup analyses will be conducted, if feasible.
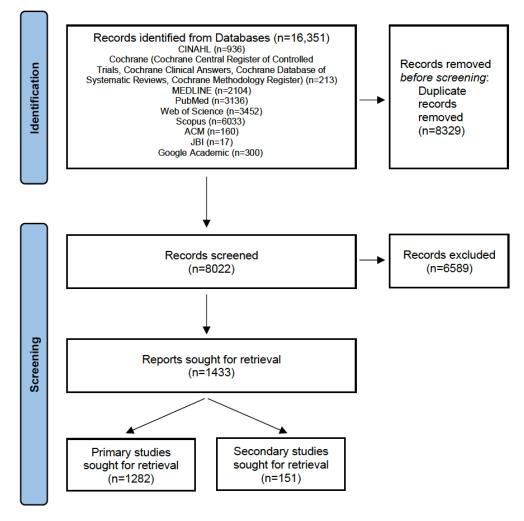
High-level maps, such as scatter charts and bubble charts, will be used to depict results and illustrate research trends and gaps. The development of the searchable database will be informed by the research questions and data extraction forms.

Results

The literature search in MEDLINE; CINAHL; Web of Science; Scopus; the Cochrane, ACM, IEEE, and JBI databases; and Google Scholar was conducted between November 11 and November 19, 2020. A total of 16,351 records were identified and exported to Rayyan. The removal of duplicates yielded 8022 records, which were subjected to screening (Figure 1).

As of April 2021, we produced an operational procedure to support screening (aided by Rayyan) to ensure consistency and reduce the amount of errors. We also drafted a glossary with operational definitions to support the study selection and data extraction procedures. This is regarded as a living document, which will be updated as our work progresses (Multimedia Appendix 3; the glossary is currently based on 5 publications [7,45-48]). Moreover, we concluded the first the stage of screening, as depicted in Figure 1.

XSL•FO

RenderX

Figure 1. A PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the literature search and selection procedures (first stage of screening) [49]. JBI: Joanna Briggs Institute.

Discussion

Our work aims to provide an overview of all relevant evidence on conversational agents for health and well-being across the life course. The evidence and gap map is a type of literature review that is particularly suited for this aim. As White et al [50] put it, "[evidence and gap maps] usually show what evidence is there, not what the evidence says." By following the same systematic approach as a systematic review, we aim to offer a wider picture of the research landscape by including both primary studies and meta-research. To the best of our knowledge, published reviews on conversational agents, such as those of Car et al [7] and Chattopadhyay et al [27], have only included primary studies. Therefore, we add to current knowledge by extending the scope of existing reviews. As for other evidence maps, the novelty of our evidence synthesis comes from its breadth (ie, summarizing all reviews and primary studies on the topic without collating effects or effect sizes), while the strength of the scoping reviews comes from their depth (ie, typically a narrower scope and direction of effect) [50]. Another aspect regarding the broader scope of our work is our definition of a conversational agent, which comprises not only chatbots but also agents with physical or web-based embodiments, such as robots or anthropomorphic web-based agents.

The number of reviews that were preliminarily uncovered by our review also suggests that it may appropriate to conduct an umbrella review of meta-research in the future. This will help those aspiring to conduct a review in this field to avoid wasteful redundancies. The number of reviews we uncovered raises the issue of the degree of overlap and the incremental value of these publications. Recently, Tugwell et al [51] elaborated on the replication of systematic reviews. In addition to direct replication, which involves repetition for verifying results, these authors put forward the concept of conceptual replication, in which a research question is broadened or narrowed to ascertain different intervention types, settings, outcomes, or study designs [51]. The data extraction of the included reviews in our work will clarify whether these reviews performed conceptual replication or undesirable repetition, which has been coined as *research waste* [51]. We envisage that the open, searchable database that will be developed in the data synthesis stage will help researchers to avoid future research waste by highlighting published reviews.

Another methodological consideration is the role of bias and its influence on the evidence map results. For instance, we addressed bias in the selection of studies through a multifold

procedure [52]. First, we detailed the research questions and the inclusion and exclusion criteria to avoid inconsistent application. Second, we developed a glossary of operational definitions to reduce discrepancies in the interpretation of key terms. Third, per the review protocol, two independent reviewers will screen and extract data, and a third reviewer will be involved when discrepancies cannot be resolved by consensus. Other procedures specified in the protocol include presenting a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for results, pilot testing the several steps of the review method (searching, selection of studies, data extraction and critical appraisal), and having workflows subsumed in the evidence map.

The evidence and gap map relies on a framework for detailing its dimensions, which will be operationalized by using row and column headings [50]. The eCCo project will tackle this issue by pursuing an integrated approach based on the interdependency of its two core activities, as follows: (1) producing an evidence map on conversational agents in health and well-being and (2) consensualizing a taxonomy on the topic. The matrix for the data extraction (Multimedia Appendix 2) was informed by the first draft of this taxonomy, and the evidence map will allow for the identification of international experts who will be involved in the consensus conference. Further, the results of the evidence map will be used to fine-tune the taxonomy draft before the draft is subjected to the scrutiny of international experts. Our research project is in line with the work of Bittner et al [53], who used an empirical-to-conceptual approach for the development of a taxonomy for conversational agents and drew upon a literature review to identify new subsets of objects. We identified a set of taxonomies in the field, albeit none were health specific [8,9]. The foci of all these taxonomies are design options that do not detail the aspects of the health intervention. Guidelines for reporting on digital health interventions [54,55] have recommended the specification of intervention components and modes of delivery (eg, specifying who delivers the intervention, who receives the intervention, how often the intervention is delivered, the intervention duration, the format of the intervention, and the context in which the intervention is delivered). These requirements were considered when first drafting the eCCo taxonomy.

None of the above-mentioned taxonomies used a scientific consensus process to standardize terminology and categories of conversational agents. This is a limitation that we are addressing via the eCCo project.

In spite of its clear strengths, the review will not be without limitations. Integral to evidence maps is the fact that study outcomes are not extracted to ascertain effects. Moreover, the fact that we will not appraise the quality of primary studies means that we cannot pinpoint research gaps for areas with high volumes of potentially poor-quality studies. Another limitation is that research gaps do not necessarily translate to research needs; when prioritizing research needs, one should consider aspects such as relevance and potential impact. Nonetheless, synthesizing evidence on health-focused conversational agents will facilitate the prioritization of strategic research by commissioners.

In addition to fostering collaboration, we envisage that the open, searchable database will also contribute to bridging the translational gap by, for instance, identifying projects with a higher technology readiness level. Such projects can more easily reach the market via partnerships with the business sector.

Acknowledgments

Authors' Contributions

MPG, LA, and EM designed the study. The remaining authors contributed to iteratively conceiving the study in more detail. MPG, HRH, MEK, CB, JB, and IBF prepared the protocol, and revisions were made by the remaining authors. HRH and MEK conducted the literature searches. All authors except MPG executed the first stage of screening.

Conflicts of Interest

None declared.

Multimedia Appendix 1

MEDLINE and Google Scholar search.
[PDF File (Adobe PDF File), 75 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Preliminary data extraction form.
[PDF File (Adobe PDF File), 85 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Operational definitions.

[PDF File (Adobe PDF File), 119 KB-Multimedia Appendix 3]

References

1. Noncommunicable diseases: mortality. World Health Organization. URL: https://www.who.int/gho/ncd/mortality_morbidity/en/ [accessed 2021-07-14]

2. Nyberg ST, Singh-Manoux A, Pentti J, Madsen IEH, Sabia S, Alfredsson L, et al. Association of healthy lifestyle with years lived without major chronic diseases. JAMA Intern Med 2020 May 01;180(5):760-768 [FREE Full text] [doi: 10.1001/jamainternmed.2020.0618] [Medline: 32250383]

3. Kelly MP, Barker M. Why is changing health-related behaviour so difficult? Public Health 2016 Jul;136:109-116 [FREE Full text] [doi: 10.1016/j.puhe.2016.03.030] [Medline: 27184821]

4. Rinaldi G, Hijazi A, Haghparast-Bidgoli H. Cost and cost-effectiveness of mHealth interventions for the prevention and control of type 2 diabetes mellitus: A systematic review. Diabetes Res Clin Pract 2020 Apr;162:108084. [doi: 10.1016/j.diabres.2020.108084] [Medline: 32061819]

5. Blueprint for a digital transformation of health and care in an ageing society. European Commission. 2017. URL: https://ec.europa.eu/digital-single-market/en/blueprint-digital-transformation-health-and-care-ageing-society [accessed 2021-01-24]

6. 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 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]

7. Car TL, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational agents in health care: Scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]

8. Diederich S, Brendel AB, Kolbe L. Towards a taxonomy of platforms for conversational agent design. 2019 Presented at: Internationale Tagung Wirtschaftsinformatik; February 2019; Siegen, Germany.

9. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Syst Appl 2019 Sep;129:56-67. [doi: 10.1016/j.eswa.2019.03.054]

10. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: Systematic review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]

11. Burden D, Savin-Baden M. What are virtual humans? In: Virtual Humans: Today and Tomorrow. Boca Raton, Florida: Chapman and Hall/CRC Press; 2019.

12. International Classification of Health Interventions (ICHI). World Health Organization. URL: https://www.who.int/standards/classifications/international-classification-of-health-interventions [accessed 2021-06-14]

13. Angelini L, Mugellini E, Abou KO, Boqué N, Del Bas J, Röcke C, et al. The Nestore e-Coach: Accompanying older adults through a personalized pathway to wellbeing. 2019 Jun Presented at: PETRA '19: The 12th PErvasive Technologies Related to Assistive Environments Conference; June 5-7, 2019; Rhodes, Greece p. 620-628. [doi: 10.1145/3316782.3322763]

14. El Kamali M, Angelini L, Caon M, Andreoni G, Khaled OA, Mugellini E. Towards the NESTORE e-coach: A tangible and embodied conversational agent for older adults. 2018 Oct Presented at: UbiComp '18: The 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing; October 8-12, 2018; Singapore, Singapore p. 1656-1663. [doi: 10.1145/3267305.3274188]

15. Félix IB, Guerreiro MP, Cavaco A, Cláudio AP, Mendes A, Balsa J, et al. Development of a complex intervention to improve adherence to antidiabetic medication in older people using an anthropomorphic virtual assistant software. Front Pharmacol 2019;10:680 [FREE Full text] [doi: 10.3389/fphar.2019.00680] [Medline: 31281256]

16. Guerreiro MP, Félix IB, Balsa J, Carmo MB, Henriques MA, Cavaco A, et al. Contribution of an intelligent virtual assistant to healthy ageing in idults with type 2 diabetes. In: Mendes D, Fonseca C, Lopes M, García-Alonso J, Murillo J, editors. Exploring the Role of ICTs in Healthy Aging. Hershey, Pennsylvania: IGI Global; 2020.

17. Harnessing the power of conversational e-coaches for health and well-being through Swiss-Portuguese Collaboration. eCCo. URL: https://ecco.esel.pt/ [accessed 2021-07-15]

18. Xing Z, Yu F, Du J, Walker JS, Paulson CB, Mani NS, et al. Conversational interfaces for health: Bibliometric analysis of grants, publications, and patents. J Med Internet Res 2019 Nov 18;21(11):e14672 [FREE Full text] [doi: 10.2196/14672] [Medline: 31738171]

19. Miake-Lye IM, Hempel S, Shanman R, Shekelle PG. What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products. Syst Rev 2016 Feb 10;5:28 [FREE Full text] [doi: 10.1186/s13643-016-0204-x] [Medline: 26864942]

20. Kamali ME, Angelini L, Caon M, Carrino F, Rocke C, Guye S, et al. Virtual coaches for older adults' wellbeing: A systematic review. IEEE Access 2020;8:101884-101902. [doi: 10.1109/access.2020.2996404]

XSL•FO

RenderX

21. Pereira J, Díaz Ó. Using health chatbots for behavior change: A mapping study. J Med Syst 2019 Apr 04;43(5):135. [doi: 10.1007/s10916-019-1237-1] [Medline: 30949846]

22. 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 Jul;64(7):456-464. [doi: 10.1177/0706743719828977] [Medline: 30897957]

23. Ma T, Chattopadhyay D, Sharifi H. Virtual humans in health-related interventions: A meta-analysis. 2019 May Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland p. 1-6. [doi: 10.1145/3290607.3312853]

24. Tropea P, Schlieter H, Sterpi I, Judica E, Gand K, Caprino M, et al. Rehabilitation, the great absentee of virtual coaching in medical care: Scoping review. J Med Internet Res 2019 Oct 01;21(10):e12805 [FREE Full text] [doi: 10.2196/12805] [Medline: 31573902]

25. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: A scoping review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]

26. Schachner T, Keller R, Wangenheim FV. Artificial intelligence-based conversational agents for chronic conditions: Systematic literature review. J Med Internet Res 2020 Sep 14;22(9):e20701 [FREE Full text] [doi: 10.2196/20701] [Medline: 32924957]

27. Chattopadhyay D, Ma T, Sharifi H, Martyn-Nemeth P. Computer-controlled virtual humans in patient-facing systems: Systematic review and meta-analysis. J Med Internet Res 2020 Jul 30;22(7):e18839 [FREE Full text] [doi: 10.2196/18839] [Medline: 32729837]

28. Bibault JE, Chaix B, Nectoux P, Pienkowsky A, Guillemasse A, Brouard B. Healthcare ex machina: Are conversational agents ready for prime time in oncology? Clin Transl Radiat Oncol 2019 Apr 04;16:55-59 [FREE Full text] [doi: 10.1016/j.ctro.2019.04.002] [Medline: 31008379]

29. 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 Jul 21;19(8):e267 [FREE Full text] [doi: 10.2196/jmir.7023] [Medline: 28784594]

30. de Cock C, Milne-Ives M, van Velthoven MH, Alturkistani A, Lam C, Meinert E. Effectiveness of conversational agents (virtual assistants) in health care: Protocol for a systematic review. JMIR Res Protoc 2020 Mar 09;9(3):e16934 [FREE Full text] [doi: 10.2196/16934] [Medline: 32149717]

31. Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: an efficient and complete method to develop literature searches. J Med Libr Assoc 2018 Oct;106(4):531-541 [FREE Full text] [doi: 10.5195/jmla.2018.283] [Medline: 30271302]

32. Bar-Ilan J. Tale of Three Databases: The Implication of Coverage Demonstrated for a Sample Query. Front Res Metr Anal 2018 Feb 20;3:6. [doi: 10.3389/frma.2018.00006]

33. Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Res Synth Methods 2020 Mar;11(2):181-217 [FREE Full text] [doi: 10.1002/jrsm.1378] [Medline: 31614060]

34. Martín-Martín A, Orduna-Malea E, Delgado López-Cózar E. Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: a multidisciplinary comparison. Scientometrics 2018 Jun 26;116(3):2175-2188. [doi: 10.1007/s11192-018-2820-9]

35. Moed HF, Bar-Ilan J, Halevi G. A new methodology for comparing Google Scholar and Scopus. J Informetr 2016 May;10(2):533-551. [doi: 10.1016/j.joi.2016.04.017]

36. Haustein S, Bowman T, Costas R. When is an article actually published? An analysis of online availability, publication, and indexation dates. arXiv Preprint posted online on May 4, 2015. [FREE Full text]

37. Lefebvre C, Glanville J, Briscoe S, Littlewood A, Marshall C, Metzendorf MI, et al. Chapter 4: Searching for and selecting studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.2. Chichester, United Kingdom: John Wiley & Sons; 2019.

38. Garousi V, Felderer M, Mäntylä MV. Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. Inf Softw Technol 2019 Feb;106:101-121. [doi: 10.1016/j.infsof.2018.09.006]

39. Haddaway NR, Collins AM, Coughlin D, Kirk S. The role of Google Scholar in evidence reviews and its applicability to grey literature searching. PLoS One 2015;10(9):e0138237. [doi: 10.1371/journal.pone.0138237] [Medline: 26379270]

40. Briscoe S, Bethel A, Rogers M. Conduct and reporting of citation searching in Cochrane systematic reviews: A cross-sectional study. Res Synth Methods 2020 Mar;11(2):169-180 [FREE Full text] [doi: 10.1002/jrsm.1355] [Medline: 31127978]

41. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]

42. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017 Dec 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]

43. Gates M, Gates A, Duarte G, Cary M, Becker M, Prediger B, et al. Quality and risk of bias appraisals of systematic reviews are inconsistent across reviewers and centers. J Clin Epidemiol 2020 Sep;125:9-15 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.04.026] [Medline: 32416337]

44. Pieper D, Puljak L, González-Lorenzo M, Minozzi S. Minor differences were found between AMSTAR 2 and ROBIS in the assessment of systematic reviews including both randomized and nonrandomized studies. J Clin Epidemiol 2019 Apr;108:26-33. [doi: 10.1016/j.jclinepi.2018.12.004] [Medline: 30543911]

45. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Trans Comput Hum Interact 2005 Jun;12(2):293-327. [doi: 10.1145/1067860.1067867]

46. de Antonio A, Aylett R, Ballin D. Intelligent virtual agents. 2001 Presented at: Third International Workshop on Intelligent Virtual Agents; September 10-11, 2001; Madrid, Spain. [doi: 10.1007/3-540-44812-8]

47. El Kamali M, Angelini L, Caon M, Khaled OA, Mugellini E, Dulack N, et al. NESTORE: Mobile chatbot and tangible vocal assistant to support older adults' wellbeing. 2020 Presented at: CUI '20: 2nd Conference on Conversational User Interfaces; July 22-24, 2020; Bilbao, Spain p. 1-3. [doi: 10.1145/3405755.3406167]

48. Poole DL, Mackworth AK. Artificial Intelligence: Foundations of Computational Agents, 2nd Edition. Cambridge, United Kingdom: Cambridge University Press; 2017.

49. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ 2021;372. [doi: 10.1136/bmj.n71] [Medline: 33782057]

50. White H, Albers B, Gaarder M, Kornør H, Littell J, Marshall Z, et al. Guidance for producing a Campbell evidence and gap map. Campbell Syst Rev 2020 Nov 19;16(4):e1125. [doi: 10.1002/cl2.1125]

51. Tugwell P, Welch VA, Karunananthan S, Maxwell LJ, Akl EA, Avey MT, et al. When to replicate systematic reviews of interventions: consensus checklist. BMJ 2020 Sep 15;370:m2864. [doi: 10.1136/bmj.m2864] [Medline: 32933948]

52. Drucker AM, Fleming P, Chan A. Research techniques made simple: Assessing risk of bias in systematic reviews. J Invest Dermatol 2016 Nov;136(11):e109-e114 [FREE Full text] [doi: 10.1016/j.jid.2016.08.021] [Medline: 27772550]

53. Bittner EAC, Oeste-Reiß S, Leimeister JM. Where is the bot in our team? Toward a taxonomy of design option combinations for conversational agents in collaborative work. 2019 Presented at: 52nd Hawaii International Conference on System Sciences; January 8-11, 2019; Maui, Hawaii, USA. [doi: 10.24251/hicss.2019.035]

54. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]

55. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

AMSTAR: A Measurement Tool to Assess Systematic Reviews
EBSCO: Elton B. Stephens Company
eCCo: Harnessing the Power of Conversational e-Coaches for Health and Well-being Through Swiss-Portuguese Collaboration
JBI: Joanna Briggs Institute
NCD: noncommunicable disease
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
ROBIS: Risk of Bias Assessment Tool for Systematic Reviews