Original Paper

Integrating Option Grid Patient Decision Aids in the Epic Electronic Health Record: Case Study at 5 Health Systems

Peter Scalia[1], MSc, PhD; Farhan Ahmad[2], BS, MBA; Danielle Schubbe[1], BA; Rachel Forcino[1], MSc, PhD; Marie-Anne Durand[1], PhD; Paul James Barr[1], MSc, PhD; Glyn Elwyn[1], PhD, MD

[1]Dartmouth College, Lebanon, NH, United States

[2]EBSCO Health, Boston, MA, United States

Corresponding Author:
Glyn Elwyn, PhD, MD
Dartmouth College
1 Medical Center Drive
Lebanon, NH, 03766
United States
Phone: 1 603 653 0806
Email: glynelwyn@gmail.com

Abstract

Background: Some researchers argue that the successful implementation of patient decision aids (PDAs) into clinical workflows depends on their integration into electronic health records (EHRs). Anecdotally, we know that EHR integration is a complex and time-consuming task; yet, the process has not been examined in detail. As part of an implementation project, we examined the work involved in integrating an encounter PDA for symptomatic uterine fibroids into Epic EHR systems.

Objective: This study aims to identify the steps and time required to integrate a PDA into the Epic EHR system and examine facilitators and barriers to the integration effort.

Methods: We conducted a case study at 5 academic medical centers in the United States. A clinical champion at each institution liaised with their Epic EHR team to initiate the integration of the uterine fibroid Option Grid PDAs into clinician-facing menus. We scheduled regular meetings with the Epic software analysts and an expert Epic technologist to discuss how best to integrate the tools into Epic for use by clinicians with patients. The meetings were then recorded and transcribed. Two researchers independently coded the transcripts and field notes before categorizing the codes and conducting a thematic analysis to identify the facilitators and barriers to EHR integration. The steps were reviewed and edited by an Epic technologist to ensure their accuracy.

Results: Integrating the uterine fibroid Option Grid PDA into clinician-facing menus required an 18-month timeline and a 6-step process, as follows: task priority negotiation with Epic software teams, security risk assessment, technical review, Epic configuration; troubleshooting, and launch. The key facilitators of the process were the clinical champions who advocated for integration at the institutional level and the presence of an experienced technologist who guided Epic software analysts during the build. Another facilitator was the use of an emerging industry standard app platform (Health Level 7 Substitutable Medical Applications and Reusable Technologies on Fast Healthcare Interoperability Resources) as a means of integrating the Option Grid into existing systems. This standard platform enabled clinicians to access the tools by using single sign-on credentials and prevented protected health information from leaving the EHR. Key barriers were the lack of control over the Option Grid product developed by EBSCO (Elton B Stephens Company) Health; the periodic Epic upgrades that can result in a pause on new software configurations; and the unforeseen software problems with Option Grid (ie, inability to print the PDA), which delayed the launch of the PDA.

Conclusions: The integration of PDAs into the Epic EHR system requires a 6-step process and an 18-month timeline. The process required support and prioritization from a clinical champion, guidance from an experienced technologist, and a willing EHR software developer team.

(J Med Internet Res 2021;23(5):e22766) doi: 10.2196/22766

KEYWORDS

shared decision making; patient decision aids; electronic health record; implementation; HL7 SMART on FHIR

Introduction

Background

Researchers have argued that the successful implementation of patient decision aids (PDAs) into clinical workflows depends on their integration into electronic health records (EHRs) [1,2]. The task of integrating third-party tools into EHRs is complex [3]. Security concerns dominate the challenge, as institutions have become reliant on EHRs to manage key operational workflows [3]. Third-party software that brings external connections and URL links to EHRs is subject to extensive scrutiny [4]. Updates to either the EHR or linked third-party products are perennial concerns, given the cost of downtime or system failure.

Many software vendors provide system-wide EHR software (eg, Epic, Cerner, or Allscripts) [5]. However, there are major differences between the same EHR product when installed at different health care institutions: this is because they are tailored to organizational and clinical preferences and integrated with other ancillary software [6]. Clinicians also differ in how and when EHRs are used with patients [7]. These uses can include showing images or test results and sending health information via the patient portal [8]. Although some common integration processes can be identified, solutions cannot be replicated from one institutional setting to others and tailoring is always required.

PDAs provide evidence-based information in a comparative format to facilitate shared decision-making, in which patients and clinicians are supported when making informed decisions together [9]. PDAs serve as catalysts to engage patients in decision-making processes and can be used before, during, and after clinical encounters [10,11]. Recent systematic reviews and meta-analyses have shown that PDAs increase knowledge about options and reduce decisional conflict, thereby helping patients make decisions that align with their preferences [12,13]. Despite improving a range of outcomes, their implementation in the clinical workflow remains a challenge [14,15]. Given the widespread adoption of EHRs and clinicians' reliance on them [16], many have presumed that integrating PDAs into EHRs will lead to their increased use in clinical practice [2,17]. However, this presumption has not yet been tested at scale.

Studies that have evaluated the integration of PDAs in EHRs have focused on measuring their use by clinicians [1,18-22], measuring their impact on patient outcomes [23-25], or user testing the tool to improve the navigation and design in the EHR system [26-29]. The integration of 2 PDAs, namely the Statin Choice and Diabetes Medication Choice tools, in the EHR at the Mayo Clinic led to their increased use [18,20,21]. Coylewright et al [1] also demonstrated an increased use and observed that adoption rates of an EHR-based *HealthDecision* tool steadily increased over an 8-year period, with a *high rate of sustained implementation after the fifth use*.

Nevertheless, we were only able to identify a few examples of PDAs being integrated into EHR systems [1,18-29]. Anecdotally, researchers and practitioners recognize that embedding PDAs in EHRs is a complex and time-consuming process; however, we could not identify the literature that described the required processes. Therefore, we lack an understanding of *how* best to integrate these tools into EHRs, and the steps required, especially given the recent development of new interoperability standards [30]. An opportunity arose to address this research gap as part of a project to implement the uterine fibroids Option Grid PDA at 5 health care institutions in the United States (Uterine Fibroids Options [UPFRONT] study) [31].

Objectives

The aims of this work are to (1) identify the steps and the time required to integrate an Option Grid PDA into the Epic EHR system and (2) examine facilitators and barriers to the integration effort. We hypothesize that some institutions will successfully integrate the Option Grid PDA into Epic as part of a multistep, time-intensive process.

Methods

Design

As part of a Patient-Centered Outcomes Research Institute (PCORI)–funded, stepped-wedge implementation trial, we asked each participating institution to integrate Option Grid PDAs into their Epic EHR systems. Despite the stepped-wedge design, we began the integration effort at all institutions almost immediately upon receiving funding to provide ample opportunity to complete the process ahead of the active implementation phase of the broader trial, which is when clinicians would be expected to use Option Grid with their patients. Successful integration was defined as the completion of changes to the Epic system that allowed clinicians to easily access an external website that provides access to both interactive and PDF versions of the uterine fibroid Option Grid PDAs [31]. Facilitators are key elements or factors that enable a successful integration. To examine the processes required, we adopted an exploratory case study design and collected data by recording relevant meetings and taking field notes [32]. We analyzed our conversations with the clinical champions, their research teams, and Epic software analysts from various departments (such as compliance, risk management, and information security) at each institution. Our case study was reported using the checklist by Rodgers et al (Multimedia Appendix 1) [33]. The Dartmouth College Committee for the Protection of Human Subjects (approval number: STUDY00031464) granted ethical approval for our study.

Settings

The implementation study took place in the following institutions: (1) Dartmouth-Hitchcock Medical Center in Lebanon, New Hampshire; (2) Barnes-Jewish Hospital in St. Louis, Missouri; (3) Montefiore Medical Center in Bronx, New York; (4) Brigham and Women's Hospital in Boston, Massachusetts; and (5) Mayo Clinic in Rochester, Minnesota. Each of these institutions had installed the Epic EHR product at different times. Our case study is based on our efforts to integrate the Option Grid into the Epic EHR system at these 5 institutions. Table 1 provides a brief description of each institution's Epic experience, expertise, and infrastructure.

Table 1. Description of each institution's Epic experience, expertise, and infrastructure.

Institution	Date of Epic adoption	Number of Epic software analysts[a]	Any previous experience with third-party software integration?	Does the clinical champion have experience with using patient decision aids?
Dartmouth-Hitchcock Medical Center	April 2011	170	Yes	Yes
Barnes-Jewish Hospital	June 2017	150	Yes	No
Montefiore Medical Center	April 2015 to June 2016	200	Yes	No
Brigham and Women's Hospital	June 2015	Unknown	Yes	Yes
Mayo Clinic	May 2018	300	Yes	No

[a]Estimated number.

Option Grid PDA

The Option Grid PDA for symptomatic uterine fibroids is part of a suite of tools developed and updated by EBSCO (Elton B Stephens Company) Health, a commercial entity that provides clinical decision support for health care organizations [34]. Our collaboration with EBSCO Health for developing and maintaining the uterine fibroids PDA came at no cost to the research effort. The uterine fibroid tool compares 7 treatment options: (1) watch and wait, (2) medicine with hormones, (3) medicine without hormones, (4) embolization, (5) endometrial ablation, (6) myomectomy, and (7) hysterectomy. The tool is available in English and Spanish for the following 2 formats: text-only and text accompanied by pictures (Picture Option Grid). For the study duration, each participating institution was granted access to the entire suite of 30 Option Grid tools.

When clinicians click the Option Grid button, they are first presented with the entire suite of PDAs. Once clinicians select the uterine fibroids Option Grid, they have the opportunity to select as many as 7 treatment options that are relevant to a particular patient. Once the PDA is generated, the clinician can use 3 features to document the options discussed in the encounter: print a PDF version, copy and paste a script that includes the options selected, or send a *permalink* to the patient so that they can view the Option Grid at their own convenience. The script is an optional feature that enables clinicians to document the use of the PDA and the options discussed in the EHR. The clinician must either use this feature or write a note in the EHR regarding the conversation that occurred with the PDA, as Option Grid does not exchange any information with the EHR. Figure 1 shows an example of an Option Grid PDA for symptomatic uterine fibroids.

Figure 1. Snapshot of the text version of the uterine fibroid Option Grid patient decision aid.

Recruitment

The 5 institutions were selected to participate in our implementation trial because of their inherent diversity, their interest in implementing PDAs, and, in some cases, their experience of practicing shared decision-making. Institutions (from inner city Bronx to rural New Hampshire) treat an ethnically diverse patient population across both urban and rural settings. At each of the 5 institutions, an obstetrics and gynecology specialist was recruited for the role of clinical

champion (site principal investigator) for their interest in using PDAs to improve health care delivery. The clinical champions contacted the Epic software analysts and explained the importance of integrating the uterine fibroids Option Grid PDA into the EHR. The clinical champion also introduced software analysts to the UPFRONT study team, which included an experienced Epic technologist employed at EBSCO Health (FA).

Data Collection

Videoconference meetings were scheduled between the UPFRONT team and the Epic software analysts at each institution. Following the introductory meeting, the Epic software analysts dictated the frequency of the meetings based on their needs for assistance or clarification from the Epic technologist. The software analysts reached out to our study team to schedule meetings. The meetings were a time to collect the work summaries that had been collected by the software analysts, identify any barriers they were facing, and discuss solutions to overcome those barriers. Each meeting was audio-recorded. The transcribed audio recordings and field notes from the meetings provided the data for analysis.

Consent

The Dartmouth College Committee for the Protection of Human Subjects waived the requirement for the written documentation of informed consent. Standard information regarding the study was provided to the participants. The study team sought verbal consent from the audio recording at the start of videoconference meetings. If verbal consent was not granted by one or more members, then the study team took notes of meeting discussions.

Analysis

For aim 1, 2 researchers (PS and DS) reviewed meeting transcripts and field notes and documented the steps taken to integrate the Option Grid into Epic. Our description of the steps was reviewed and edited by the Epic technologist (FA) and modifications were made if required. To determine the amount of time required for the integration, we counted the number of months from the original email to the clinical champions to gauge their interest in a possible integration effort to the moment the Option Grid was launched in the site's Epic environment. For aim 2, we conducted an inductive thematic analysis of the transcripts and field notes. Two researchers (PS and DS) independently coded a sample of transcript pages and then discussed and agreed on a codebook. The finalized codebook (Multimedia Appendix 2) was applied to all the data to highlight the facilitators and barriers to the integration effort. Codes were grouped into different code categories, which were revised and discussed by PS and DS to determine the themes. Coding disagreements were resolved by a third researcher (GE).

Results

Overview

A total of 27 meetings were held. These included 25 videoconferences (Zoom, Webex [Cisco], or Skype [Microsoft Corporation]) and 2 telephone meetings. Of the 27 meetings, we were able to record 22 (81%) meetings and collated notes for the remaining 5 (19%). The bidirectional exchange of information between the Epic software analysts and the study team (including the Epic technologist) during the meetings at each institution yielded a total of 183 transcripts or field note pages (total word count: 91,336; see Table 2 for details).

Table 2. Details of the 27 meetings that informed the steps to integrate the patient decision aid into the electronic health record and the facilitators and barriers to the integration effort.

Institution and meeting date	Platform	Recorded? (yes or no)	Meeting duration (min:sec)	Meeting personnel
Dartmouth-Hitchcock Medical Center				
January 9, 2019	Zoom	Yes	40:03	3 UPFRONT[a] study members, 4 EBSCO[b] Health members, and 2 clinical informatics members and clinicians
February 6, 2019	Phone	Yes	44:50	2 UPFRONT study members, 1 clinical champion, 1 Epic technologist, 1 EBSCO Health member, 1 clinical informatics member and clinician, and 2 clinicians
July 25, 2019	Webex	Yes	23:39	3 UPFRONT study members and 1 Epic technologist
August 9, 2019	Zoom	Yes	5:49	2 UPFRONT study members, 1 Epic technologist, and 1 clinical informatics member and clinician
Barnes-Jewish Hospital				
February 4, 2019	Zoom	Yes	50:24	3 UPFRONT study members, 1 clinical champion, 1 Epic technologist, 2 EBSCO Health members, and 1 research assistant
May 23, 2019	Zoom	Yes	27:25	3 UPFRONT study members, 1 Epic technologist, and 1 Epic software analyst
June 6, 2019	Zoom	Yes	35:07	3 UPFRONT study members, 2 Epic software analysts, and 1 Epic technologist
February 3, 2020	Webex	Yes	MDNR[c]	1 UPFRONT study member, 1 Epic software analyst, 2 clinicians, and 18 ambulatory operations members
February 11, 2020	Webex	Yes	MDNR	1 UPFRONT study member, 1 Epic technologist, 1 Epic software analyst, and 1 interfaces team member
March 3, 2020	Webex	Yes	27:35	1 UPFRONT study member, 1 Epic technologist, and 1 Epic software analyst
Montefiore Medical Center				
January 22, 2019	Zoom	Yes	7:45	2 clinical champions, 1 UPFRONT study member, 1 Epic software analyst, and 3 research assistants
January 30, 2019	Zoom	Yes	27:38	2 UPFRONT study members, 2 EBSCO Health members, and 4 Epic software analysts
January 14, 2020	Skype	No	__[d]	1 clinical champion, 2 UPFRONT study members, 1 Epic technologist, 2 Epic software analysts, and 2 research assistants
Brigham and Women's Hospital				
January 25, 2019	Zoom	Yes	21:59	2 UPFRONT study members and 1 partners operations member
February 18, 2019	Zoom	Yes	33:28	2 UPFRONT study members and 2 Epic software analysts
March 27, 2019	Zoom	Yes	15:45	2 UPFRONT study members and 2 Epic software analysts
May 22, 2019	Zoom	Yes	37:08	3 UPFRONT study members, 1 Epic technologist, 5 Epic software analysts, and 1 research assistant
July 29, 2019	Zoom	Yes	18:42	2 UPFRONT study members, 1 Epic technologist, 1 partners operations member, and 2 research assistants
January 9, 2020	Webex	Yes	MDNR	1 UPFRONT study member, 1 Epic technologist, and 2 Epic software analysts
March 30, 2020	Webex	Yes	MDNR	1 clinical champion, 1 UPFRONT study member, 1 Epic technologist, and 2 Epic software analysts
April 24, 2020	Phone	No	—	1 UPFRONT study member, 1 Epic technologist, 1 research assistant, and 2 information security analysts
May 28, 2020	Webex	No	—	1 Clinical champion, 1 Epic technologist, 2 Epic software analysts, and 1 research assistant
June 5, 2020	Webex	No	—	1 UPFRONT study member, 1 Epic technologist, 2 Epic software analysts, and 1 EBSCO Health member

Institution and meeting date	Platform	Recorded? (yes or no)	Meeting duration (min:sec)	Meeting personnel
June 23, 2020	Webex	No	—	1 UPFRONT study member, 1 Epic technologist, and 2 Epic software analysts
Mayo Clinic				
January 23, 2019	Webex	Yes	47:59	1 Clinical champion, 1 UPFRONT study member, and 3 Epic software analysts
July 24, 2019	Zoom	Yes	47:33	1 clinical champion, 3 UPFRONT study members, 1 Epic technologist, 3 Epic software analysts, and 1 research assistant
January 29, 2020	Webex	Yes	MDNR	1 UPFRONT study member, 1 Epic technologist, and 3 Epic software analysts

[a]UPFRONT: Uterine Fibroids Options.

[b]EBSCO: Elton B Stephens Company.

[c]MDNR: meeting duration not recorded.

[d]Not available. Participants did not agree to be recorded.

Aim 1: The Steps Taken to Integrate the PDA Into Epic

We were able to describe the process of integrating the Option Grid PDAs into Epic by identifying 6 common process steps across the institutions (Textboxes 1 and 2).

The timeline for completing the 6 steps varied across institutions, but overall, up to 18 months (January 2019 to June 2020) was required to integrate the Option Grid PDAs into the Epic EHR system. The timeline began in January 2019, which was when all clinical champions received an email to gauge interest in integrating Option Grid into a clinician-facing menu in their Epic system. Although work began to place a button in an agreed location nominated by clinicians at each institution, a policy decision was made by EBSCO Health in August 2019 to use Substitutable Medical Applications and Reusable Technologies (SMART) and Fast Healthcare Interoperability Resources (FHIR) standards and list the application on Epic's App Orchard to simplify the setup and maintenance of the tool throughout the study period. Owing to the COVID-19 pandemic, integration efforts at 2 institutions were paused to redirect resources to other more pressing initiatives. One institution was able to resume the effort, whereas the other institution had competing priorities and was forced to furlough some personnel involved in the integration process. Thus, we were only able to complete the integration of Option Grid into Epic at 4 of the 5 sites. Figure 2 details the integration timeline following the policy decision, beginning with the security risk assessment (Step 2) and ending with the Option Grid launch (Step 6).

Textbox 1. Steps to integrate the Option Grid patient decision aids into the electronic health record.

Step 1: Negotiating task priority with Epic software teams

- The clinical champions at each site requested the changes into their Epic systems, establishing the clinical benefit of providing easy access to the Option Grid website in a clinician-facing menu location. Epic teams had to reprioritize their tasks, given existing work schedules. In one setting, agreement to reprioritize required negotiation and financial support

Step 2: Security risk assessment

- Each institution had different security risk assessment processes, with each requiring departmental approval. Typically, 3 levels of security checks were required, related to the changes in outpatient processes or menus, information flows and dependencies, and communication with third-party tools. Epic has different modules (ie, outpatient, inpatient, and research) and each has its own operations and approval groups. The Ambulatory Operations group is responsible for reviewing requests related to general outpatient workflows (Barnes-Jewish Hospital and Montefiore Medical Center). In the case of Barnes-Jewish Hospital, the integration had to also be approved by the Interfaces Operations group, which manages requests related to any information that gets moved in and out of Epic via interfaces, and the Infrastructure team, which manages the App Orchard. The Mayo Clinic also had 3 levels of security: (1) Security, Privacy, Architecture and Data Assurance reviews new technology that will be integrated into Epic; (2) Clinical Decision Support reviews electronic health record (EHR) change requests that have endorsement from a clinical or practice committee and includes some form of clinical decision support (ie, patient decision aids [PDAs]); and (3) the Obstetrics and Gynecology specialty group, which reviews any new process, procedure, or app that will be integrated into Epic from the clinical perspective. The 4 levels of security at Brigham and Women's Hospital include (1) the clinical vetting that reviews the study context with 2 clinicians; (2) the technical feasibility of the project evaluated by an Epic team leader, (3) technical assessment (see Step 3 for details); and (4) the security risk assessment that represents an internal process to review the application being integrated into Epic. Security review was considered unnecessary at Dartmouth-Hitchcock because the Option Grid tools were merged with an existing decision support product—HealthDecision. Analysts were particularly focused on data exchange requirements between the Option Grid app and Epic. The lack of protected health information (PHI) transfer was important. Using Substitutable Medical Applications and Reusable Technologies on Fast Healthcare Interoperability Resources to allow approval for the synchronization of the private Option Grid App Orchard app with the institution's Epic environment (Step 4 for details) standardized the process across all 5 sites

Step 3: Technical review

- The Epic software analysts determined the number of personnel and time required for the overall software changes. Once the level of effort, allocation of tasks, and timelines were established, the software analysts were ready for the build

Step 4: Epic configuration

- Before commencing the build, all clinical champions indicated their preference for placing the Option Grid button in the patient's chart under the More menu in their toolbar at the top. Some clinical champions preferred that the button be accessible to the entire institution, whereas others preferred a restricted access to their obstetrics and gynecology department. Our study team's Epic technologist then developed and shared a build guide with each institution. The guide outlined the study objectives, the Epic-specific configurations required, and described how the Option Grid would not require access to PHI. The Epic software analysts configured access to the Option Grid PDA using Health Level 7 Substitutable Medical Applications and Reusable Technologies on Fast Healthcare Interoperability Resources (see Textbox 2 for further details). The guide contained 4 steps:
    - Accessing the app at the Epic App Orchard
    - Enabling synchronization of the Option Grid PDA on the App Orchard with the institution's Epic environment
    - Establishing a test environment before launch in a production environment
    - Requesting whitelisting of the Option Grid domains so that the app could be accessed within Epic's EHR menus, tasks, and options. Epic uses the term Hyperspace to describe this view of the software. Launching Option Grid within Hyperspace allows clinicians to have an easy access to the PDAs, without having to navigate to an external website

Step 5: Troubleshooting

- The software analysts ensured that menu locations, access requests, and user identification were all functioning as planned. This step represented a final check to ensure that all the Option Grid features were operational

Step 6: Launch

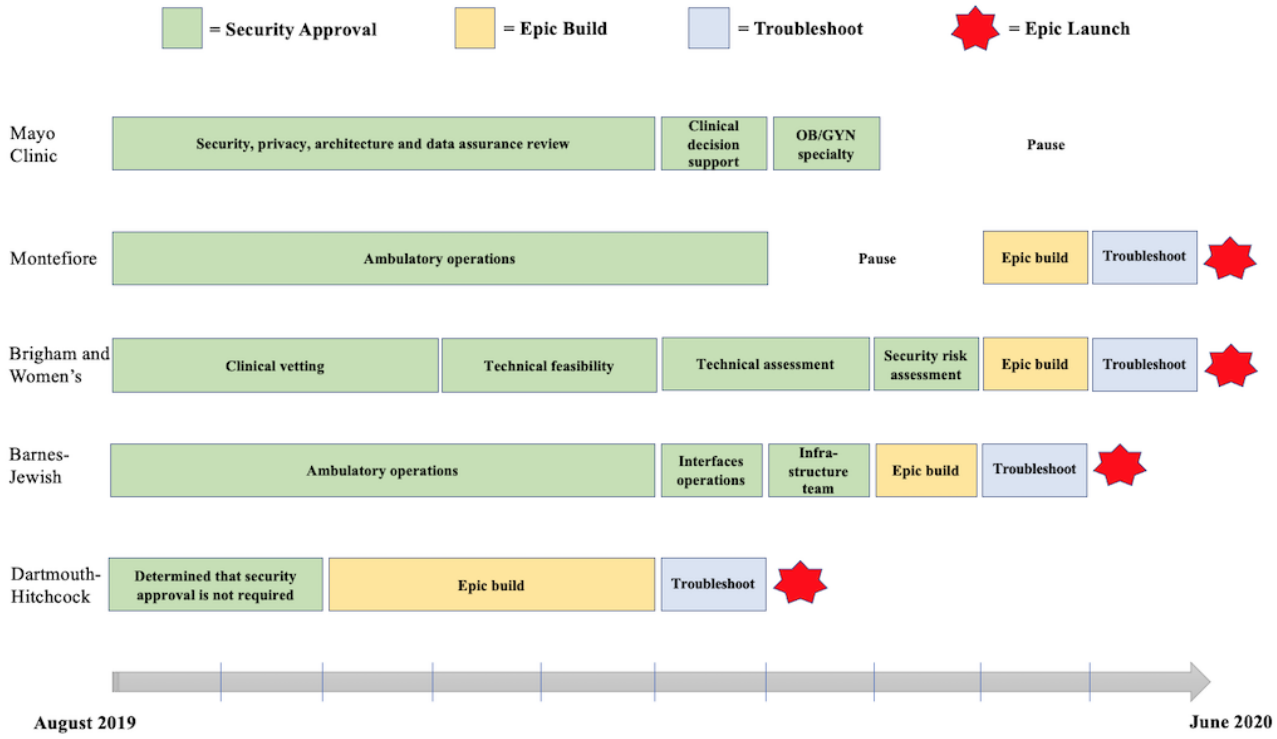
- After troubleshooting, the new configuration was migrated to the production environment and the Option Grid was launched. This means that clinicians could access the Option Grid button and be directed to the Option Grid website where they can generate a PDA for their patients

Textbox 2. Health Level 7 Substitutable Medical Applications and Reusable Technologies on Fast Healthcare Interoperability Resources authentication.

SMART on FHIR

- Substitutable Medical Applications and Reusable Technologies (SMART) is an "open, standards-based platform that enables innovators to create applications that seamlessly and securely run across the healthcare system" [35]. Health Level 7 (HL7) is an "industry organization that develops standards for the exchange, integration, sharing and retrieval of EHR information" [36]. HL7 adopted the OAuth2-based SMART App Launch framework as a core interoperability standard. HL7 also developed the Fast Healthcare Interoperability Resources (FHIR) standard to ensure *interoperability, extensibility, and speed* while searching for information across clinical applications [37]. SMART, along with FHIR—collectively referred to as SMART on FHIR—connects third-party apps to Epic, enabling them to reliably and securely launch in Epic's Hyperspace desktop client [38]. We used SMART on FHIR authentication for the following 3 reasons:
  - Improved analytics: allows the tracking of app use so that we can determine the number of eligible patients who received the uterine fibroid Option Grid patient decision aid (PDA)
  - Improved control for data access: SMART on FHIR allows a better control of the information shared with third-party apps
  - Uses existing authorizations: using FHIR allows clinician access to the Option Grid PDAs in their existing user interface (Epic's Hyperspace)
- To leverage SMART on FHIR, Epic requires apps to participate in its App Orchard app store. For the UPFRONT (Uterine Fibroids Options) study, EBSCO (Elton B Stephens Company) chose to list Option Grid as a private app on the Epic App Orchard and allowed access by the 5 participating institutions. This also simplified the setup and maintenance of the PDA app during the study period.

Figure 2. Timeline of the electronic health record integration of the Option Grid patient decision aid. OB/GYN: obstetrics and gynecology.



Aim 2: Thematic Analysis of Facilitators and Barriers to EHR Integration

We identified 4 integration facilitators (Textbox 3):

Textbox 3. Facilitators of electronic health record integration.

Facilitators

1. Clinical advocacy: presence of a clinical champion at each institution
2. Electronic health record expertise: presence of an Epic technologist with experience in building apps in Epic
3. Standardization of process: use of Substitutable Medical Applications and Reusable Technologies on Fast Healthcare Interoperability Resources standards
4. Avoidance of protected health information (PHI) data transfer: no exchange of PHI between Option Grid and the institution's local Epic environment

Clinical Advocacy

The clinical champions provided the required professional arguments for integration and served as a gateway to the Epic software analysts and lobbied for prioritization of the task. For instance, the principal investigator at Montefiore Medical Center advocated for the Epic software analysts to prioritize the integration effort:

> *We are getting it prioritized by the Epic work group with a goal for the summer.* [Principal investigator at the Montefiore Medical Center]

EHR Expertise

Having an EHR expert on our study team with a direct experience of Epic EHR environments and of the app's requirements at EBSCO was a major facilitator. He developed a *build guide*, which facilitated the integration of Option Grid PDAs in Epic at the 5 settings. Throughout the meetings, the EHR expert supported the software analysts by answering questions, clarifying points of confusion, and providing solutions for any technical issues that arose throughout the process. The following is an example of an exchange between the Epic expert and a software analyst at Barnes-Jewish Hospital during the troubleshooting step of the integration process:

> *The environment is listed in our App Orchard listing and has the test client ID. Can we test? [launch URL provided] The tokens in the OAuth context values are the same.* [Epic technologist]
>
> *Thank you—I updated the URL item in the record and tested it again. That seemed to work! I am able to open the Uterine Fibroids Option Grid, copy and paste information, open a PDF, and print the PDF.* [Barnes-Jewish hospital software analyst]

Standardization of Process

First, the use of SMART on FHIR standards provided a helpful and officially sanctioned way to better control the information being shared with a third party such as the Option Grid PDA. As described in Textbox 2, the process provided clinicians with single sign-on credentials and enabled us to track the number of times each clinician generated or accessed a uterine fibroid PDA. The Epic technologist also indicated that Epic recommends SMART on FHIR over alternative launch methods such as active guidelines and other URL-based methods:

> *The reason we went with SMART authentication is because Epic specifically advised us to do so for newer implementations.* [Epic technologist]

Avoidance of Protected Health Information Data Transfer

Placing the Option Grid PDA as a private app in Epic's App Orchard restricted Option Grid from retrieving protected health information (PHI). Option Grid instructed Epic to block certain features (ie, incoming application programming interface), so their organization would be unable to attain PHI. This arrangement eased the security concerns at each institution. For example, a software analyst at the Mayo Clinic informed us that it would be easier to obtain approval from the security team if PHI was not leaving their Epic environment:

> *When we bring in new technology it has to go to Security, Privacy, Architecture and Data Assurance (SPAD) review, but if there is no PHI it will get blessed faster.* [Mayo software analyst]

We identified 3 themes that represented barriers to the integration effort (Textbox 4).

Textbox 4. Barriers of electronic health record integration.

Barriers
1. Commercial third-party autonomy: lack of control over the Option Grid patient decision aid product owned by EBSCO Health
2. Electronic health record updates and maintenance: periodic Epic upgrades causing some delays and functionality issues
3. Unforeseen software problems: found while troubleshooting the app leading to minor delays to launch

Commercial Third-Party Autonomy

Although our collaboration with EBSCO was an overall facilitator, it also presented a barrier. EBSCO owns the Option Grid product, and as researchers, we have no influence on product development. Future product development may require additional integration efforts during the life of the study and remain a risk area. For instance, the policy decision in August 2019 to change the integration strategy forced software analysts to adapt and use a different configuration than originally planned. The following is a part of the study team's communication to each institution, explaining the shift in integration strategy:

> *EBSCO has been working on enhancing Option Grid and its integration with Epic. This process will utilize SMART and FHIR standards. As a result, we are requesting to delay the integration build in your Epic environment.* [UPFRONT study team]

EHR Updates and Maintenance

Some institutions have regular Epic upgrades and either refrain from integrating apps during those upgrades or impose a freeze on all new software configurations. The software analysts did their best to plan ahead and sidestep this barrier at the final step. The following are 2 quotes that highlight reluctance to launch the button during an upgrade:

> *Some customers have code freezes that could last more than a month when they have Epic upgrades.* [EHR technologist]
>
> *We generally don't push out new applications during an Epic upgrade.* [Brigham software analyst]

Unforeseen Software Problems

We experienced unforeseen software-related problems during the troubleshooting steps. In some settings, clinicians were unable to print PDF versions of the tool or copy and paste the script documenting the uterine fibroid options discussed in the patient's record. These software issues were conveyed to our EHR technologist:

> *We confirmed that the copy/paste does not work when launching the application.* [Dartmouth software analyst]
>
> *The PDF button did not work. I was still able to navigate elsewhere on the page, but the PDF button was non-responsive.* [Brigham software analyst]

These issues were quickly resolved through collaboration between the software analysts and the Epic technologist and did not significantly delay the launch of the Option Grid in the site's Epic environments.

Discussion

Principal Findings

The integration of Option Grid PDAs into an EHR such as Epic requires clinical advocacy, a standardized process that avoids using PHI, and expertise to guide the process. Without the support of a clinical champion in each setting, we would not have been able to initiate the process of PDA integration. At the core of the work are issues of security and reassuring the organization that data transfers will not breach security protocols. SMART on FHIR addresses the data security requirements by allowing for a better control of the information being shared with a third party such as Option Grid. The availability of an EHR expert on our study team provided the necessary guidance and reassurance to the existing Epic teams. With all these components and facilitators present, the integration process took up to 18 months to achieve. Barriers were the lack of control over the Option Grid product, EHR updates and maintenance, and unforeseen software problems that caused delays and functionality issues.

Strengths and Limitations

Our use of a case study method to elicit a real-world, in-depth understanding of the steps required to integrate PDAs into Epic, and the associated barriers and facilitators, is novel and provides new insights. However, this study has limitations. First, we were only able to examine the Epic EHR system, so we do not know if our description of the integration process steps applies to other systems. Second, the Option Grid PDAs do not require the use of PHI data. Some PDAs being developed require the exchange of PHI, which we suspect would prolong integration timelines at many institutions. Third, all the institutions were large academic medical centers with expertise in configuring the Epic EHR system, so we do not know if our findings are applicable to smaller clinical practices that lack such capability. Fourth, we did not conduct a thematic analysis to address our first aim. However, despite the absence of a thematic analysis, we feel like a review of meeting transcripts, with oversight from an Epic technologist, represents the appropriate method to determine the steps to integrate the Option Grid PDA into Epic. Furthermore, it is not known whether EBSCO Health would provide an Epic technologist to other customers or organizations aiming to integrate Option Grid in their Epic. Finally, because of the COVID-19 pandemic, we were only able to integrate the Option Grid PDAs at 4 of the 5 institutions. One institution had to redirect resources to other initiatives and were faced with staffing limitations, which led to a pause in the integration effort. We did not include this as a barrier, considering that under normal circumstances, the institution would be positioned to complete the integration of Option Grid in their Epic.

Results in Context

To the best of our knowledge, this is the first study to describe SMART on FHIR standards to integrate third-party PDAs, such as Option Grid, into an EHR system. Our results address an important gap outlined by a recent feasibility study that integrated a third-party prostate cancer screening PDA app into the EHR [3]. The authors of that study recognized the potential of SMART on FHIR to standardize secure data exchange and enable integration across a variety of EHRs [3]. However, research has focused on the interoperability of FHIR standards. For instance, a recent review showed how FHIR moved clinical information (medical images and quality metrics) found on different platforms in the EHR into a single platform to streamline the workflow of radiologists [37]. Similarly, another system has used FHIR standards to collect data from multiple sources in the EHR, automate analyses of laboratory test results, and generate easy-to-read reports for patients and their clinicians [39].

For aim 2, a key facilitator of Option Grid PDA integration into Epic systems was the presence of a clinical champion. This aligns with the results of an effort to integrate an EHR-based PDA in the emergency department for concussion and brain injury decisions [40]. They reported the critical need to engage clinicians and other information technology stakeholders [40]. In our case, the clinical champion served as an intermediary between the study team and the Epic software analysts, facilitated prioritization, and identified the EHR menu button location to ensure visibility. Clinicians' input in the integration process is reported to potentially cause the sustained use of the tools in practice [41,42] and is key to an integration effort, regardless of the format or mode of delivery [14,43-45].

Implications

Integrating third-party software into EHR systems requires a clinical champion to advocate for the task at the institutional level and an EHR expert who can guide software analyst teams throughout the process. From a policy perspective, implementing SMART on FHIR-compatible servers, which has been done at Duke Medical Center, can improve interoperability and the seamless integration of patient-facing apps [46]. However, the technologies for standardizing the integration of various types of apps, such as Option Grid, do not necessarily mean that they will be used in clinical practice. Integration, though difficult to achieve, seems to be the first step to ensure that clinicians have access to such tools. Providing an integration guide for other organizations to follow and identifying the barriers and facilitators of the process will enable more tools to be integrated into workflows. However, access alone might not lead to use,

as observed by others [21]. Future work should include usability, acceptability, or health technology assessments to further evaluate how to trigger clinicians to access and use embedded PDAs [22].

Conclusions

Integrating the uterine fibroid Option Grid PDA into clinician-facing menus in Epic was an approximately 18-month process, facilitated by a clinical champion who lobbied for the prioritization of the effort at the institutional level, and an EHR expert who guided the Epic software analysts throughout the study. The use of Health Level 7 SMART on FHIR standardized the integration effort, provided clinicians with single sign-on credentials, and more importantly blocked the exchange of PHI between Epic and Option Grid PDAs. Whether integration leads to patient use remains an open question.

Acknowledgments

The authors would like to thank the clinical champions at each institution for spearheading the EHR integration effort and serving as the gateway to the Epic software analysts: Dr Erika Banks, Dr Raymond M Anchan, Dr Shannon K Laughlin-Tommaso, Dr Tessa Madden, and Dr Tina C Foster. They also appreciate the time and effort of every Epic software analyst they encountered in their journey. They want to thank Dr Jon Keevil at EBSCO for working with site personnel to synchronize the private Option Grid App Orchard app with their Epic environment. Research reported in this study was funded by the PCORI award (grant SDM-2017C2-8507). The statements presented in this study are solely the responsibility of the authors and do not necessarily represent the views of the PCORI, its Board of Governors, or the Methodology Committee.

Conflicts of Interest

GE has edited and published books that provide royalties on sales by publishers, including Shared Decision Making (Oxford University Press) and Groups (Radcliffe Press). GE's academic interests focus on shared decision making and coproduction. He owns copyright in measures of shared decision making and care integration, namely collaboRATE, integRATE (measure of care integration), consideRATE (patient experience of care in serious illness), coopeRATE (measure of goal setting), toleRATE (clinician attitude to shared decision making), Observer OPTION-5, and Observer OPTION-12 (observer measures of shared decision making). In the past, he provided consultancy for organizations, including (1) Emmi Solutions LLC (limited liability company) who developed patient decision support tools; (2) the National Quality Forum on the certification of decision support tools; (3) Washington State Health Department on the certification of decision support tools; and (4) SciMentum LLC, Amsterdam (workshops for shared decision making). He is the founder and director of &think LLC, which owns the registered trademark for Option Grid PDAs, and founder and director of SHARPNETWORK LLC, a provider of training for shared decision making. He provides advice in the domain of shared decision making and PDAs to (1) Access Community Health Network, Chicago (Adviser to Federally Qualified Medical Centers); (2) EBSCO Health for Option Grids PDAs (consultant); (3) Bind On Demand Health Insurance (consultant); (4) PatientWisdom Inc (Adviser); and (5) Abridge AI (artificial intelligence) Inc (Chief Clinical Research Scientist).

MAD is a consultant to Access Community Health Network. Together with GE, she developed the Option Grid PDAs, which are licensed to EBSCO Health. She receives consulting income from EBSCO Health and may receive royalties in the future.

FA is the Vice President of Clinical Decision Support Technology at EBSCO Health with existing and future applications on the Epic App Orchard and a previous Epic employee.

No other authors have conflicts of interest to declare.

Multimedia Appendix 1

The guidelines for the reporting of organizational case studies.
[DOCX File , 13 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Code hierarchy used to identify the facilitators and barriers of the electronic health record integration effort.
[DOCX File , 20 KB-Multimedia Appendix 2]

References

1. Coylewright M, Keevil JG, Xu K, Dodge SE, Frosch D, Field ME. Pragmatic Study of Clinician Use of a Personalized Patient Decision Aid Integrated into the Electronic Health Record: An 8-Year Experience. Telemed J E Health 2020 May;26(5):597-602. [doi: 10.1089/tmj.2019.0112] [Medline: 31381477]
2. Lenert L, Dunlea R, Del FG, Hall LK. A model to support shared decision making in electronic health records systems. Med Decis Making 2014 Nov;34(8):987-995 [FREE Full text] [doi: 10.1177/0272989X14550102] [Medline: 25224366]
3. Day FC, Pourhomayoun M, Keeves D, Lees AF, Sarrafzadeh M, Bell D, et al. Feasibility study of an EHR-integrated mobile shared decision making application. Int J Med Inform 2019 Dec;124:24-30. [doi: 10.1016/j.ijmedinf.2019.01.008] [Medline: 30784423]

4. Tejero A, de LTI. Advances and current state of the security and privacy in electronic health records: survey from a social perspective. J Med Syst 2012 Oct;36(5):3019-3027. [doi: 10.1007/s10916-011-9779-x] [Medline: 21947856]
5. Ratwani RM, Savage E, Will A, Arnold R, Khairat S, Miller K, et al. A usability and safety analysis of electronic health records: a multi-center study. J Am Med Inform Assoc 2018 Sep 01;25(9):1197-1201. [doi: 10.1093/jamia/ocy088] [Medline: 29982549]
6. Ben-Assuli O. Electronic health records, adoption, quality of care, legal and privacy issues and their implementation in emergency departments. Health Policy 2015 Mar;119(3):287-297. [doi: 10.1016/j.healthpol.2014.11.014] [Medline: 25483873]
7. McGinn T. Putting Meaning into Meaningful Use: A Roadmap to Successful Integration of Evidence at the Point of Care. JMIR Med Inform 2016;4(2):e16 [FREE Full text] [doi: 10.2196/medinform.4553] [Medline: 27199223]
8. Zhao JY, Song B, Anand E, Schwartz D, Panesar M, Jackson GP, et al. Barriers, Facilitators, and Solutions to Optimal Patient Portal and Personal Health Record Use: A Systematic Review of the Literature. AMIA Annu Symp Proc 2017;2017:1913-1922 [FREE Full text] [Medline: 29854263]
9. Elwyn G, Lloyd A, Joseph-Williams N, Cording E, Thomson R, Durand M, et al. Option Grids: shared decision making made easier. Patient Educ Couns 2013 Feb;90(2):207-212. [doi: 10.1016/j.pec.2012.06.036] [Medline: 22854227]
10. Tsulukidze M, Grande SW, Gionfriddo MR. Assessing Option Grid® practicability and feasibility for facilitating shared decision making: An exploratory study. Patient Educ Couns 2015 Jul;98(7):871-877. [doi: 10.1016/j.pec.2015.03.013] [Medline: 25841546]
11. Seal RP, Kynaston J, Elwyn G, Smith PEM. Using an Option Grid in shared decision making. Pract Neurol 2014 Mar;14(1):54-56. [doi: 10.1136/practneurol-2013-000666] [Medline: 24151338]
12. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Dec 12;4:CD001431. [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
13. Scalia P, Durand M, Berkowitz JL, Ramesh NP, Faber MJ, Kremer JAM, et al. The impact and utility of encounter patient decision aids: Systematic review, meta-analysis and narrative synthesis. Patient Educ Couns 2019 May;102(5):817-841. [doi: 10.1016/j.pec.2018.12.020] [Medline: 30612829]
14. Elwyn G, Scholl I, Tietbohl C, Mann M, Edwards AGK, Clay C, et al. "Many miles to go …": a systematic review of the implementation of patient decision support interventions into routine clinical practice. BMC Med Inform Decis Mak 2013;13 Suppl 2:S14 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S14] [Medline: 24625083]
15. Dobler CC, Sanchez M, Gionfriddo MR, Alvarez-Villalobos NA, Singh Ospina N, Spencer-Bonilla G, et al. Impact of decision aids used during clinical encounters on clinician outcomes and consultation length: a systematic review. BMJ Qual Saf 2019 Jun;28(6):499-510 [FREE Full text] [doi: 10.1136/bmjqs-2018-008022] [Medline: 30301874]
16. Furukawa MF, King J, Patel V, Hsiao C, Adler-Milstein J, Jha AK. Despite substantial progress In EHR adoption, health information exchange and patient engagement remain low in office settings. Health Aff (Millwood) 2014 Sep;33(9):1672-1679. [doi: 10.1377/hlthaff.2014.0445] [Medline: 25104827]
17. Ankolekar A, Dekker A, Fijten R, Berlanga A. The Benefits and Challenges of Using Patient Decision Aids to Support Shared Decision Making in Health Care. JCO Clin Cancer Inform 2018 Dec;2:1-10 [FREE Full text] [doi: 10.1200/CCI.18.00013] [Medline: 30652607]
18. Inselman J, Branda M, Castaneda-Guarderas A, Gionfriddo MR, Zeballos-Palacios CL, Morris MM, et al. Uptake and Documentation of the Use of an Encounter Decision Aid in Usual Practice: A Retrospective Analysis of the Use of the Statin/Aspirin Choice Decision Aid. Med Decis Making 2016 May;36(4):557-561. [doi: 10.1177/0272989X15618175] [Medline: 26721257]
19. Jouni H, Haddad RA, Marroush TS, Brown S, Kruisselbrink TM, Austin EE, et al. Shared decision-making following disclosure of coronary heart disease genetic risk: results from a randomized clinical trial. J Investig Med 2017 Mar;65(3):681-688 [FREE Full text] [doi: 10.1136/jim-2016-000318] [Medline: 27993947]
20. Ye S, Leppin AL, Chan AY, Chang N, Moise N, Poghosyan L, et al. An Informatics Approach to Implement Support for Shared Decision Making for Primary Prevention Statin Therapy. MDM Policy Pract 2018;3(1):2381468318777752 [FREE Full text] [doi: 10.1177/2381468318777752] [Medline: 30288449]
21. Ballard AY, Kessler M, Scheitel M, Montori VM, Chaudhry R. Exploring differences in the use of the statin choice decision aid and diabetes medication choice decision aid in primary care. BMC Med Inform Decis Mak 2017 Aug 10;17(1):118 [FREE Full text] [doi: 10.1186/s12911-017-0514-5] [Medline: 28797295]
22. Percefull J, Butler J. Improving mammography through effective screening, brief intervention, and referral to treatment at a rural health center. J Am Assoc Nurse Pract 2020 Jan 16:324-330. [doi: 10.1097/JXX.0000000000000356] [Medline: 31972786]
23. Denig P, Schuling J, Haaijer-Ruskamp F, Voorham J. Effects of a patient oriented decision aid for prioritising treatment goals in diabetes: pragmatic randomised controlled trial. BMJ 2014 Sep 25;349:g5651 [FREE Full text] [doi: 10.1136/bmj.g5651] [Medline: 25255799]

24. Schott SL, Xu K, Berkowitz J, Petersen C, Saunders C, Sobti N, et al. Timing of electronic health record integrated decision aid (IDEA) for stroke prevention in atrial fibrillation matters. Journal of the American College of Cardiology 2019 Mar;73(9):3005. [doi: 10.1016/s0735-1097(19)33611-3]
25. Ajeesh S, Luis R. A Comprehensive Electronic Health Record Based Patient Navigation Module Including Technology Driven Colorectal Cancer Outreach and Education. J Cancer Educ 2018 Jun;33(3):627-633. [doi: 10.1007/s13187-017-1184-9] [Medline: 28188568]
26. Wan Q, Makeham M, Zwar NA, Petche S. Qualitative evaluation of a diabetes electronic decision support tool: views of users. BMC Med Inform Decis Mak 2012 Jul 03;12:61 [FREE Full text] [doi: 10.1186/1472-6947-12-61] [Medline: 22759239]
27. Finkelstein J, Wood J, Crew K, Kukafka R. Introducing a comprehensive informatics framework to promote breast Cancer risk assessment and chemoprevention in the primary care setting. AMIA Summits on Translational Science Proceedings Jul 2017 Jul 26:58-67. [doi: 10.1109/ichi.2016.41] [Medline: 28815107]
28. Richardson S, Feldstein D, McGinn T, Park LS, Khan S, Hess R, et al. Live Usability Testing of Two Complex Clinical Decision Support Tools: Observational Study. JMIR Hum Factors 2019 May 15;6(2):e12471 [FREE Full text] [doi: 10.2196/12471] [Medline: 30985283]
29. Jenssen BP, Shelov ED, Bonafide CP, Bernstein SL, Fiks AG, Bryant-Stephens T. Clinical Decision Support Tool for Parental Tobacco Treatment in Hospitalized Children. Appl Clin Inform 2016;7(2):399-411 [FREE Full text] [doi: 10.4338/aci-2015-12-ra-0169] [Medline: 27437049]
30. Kuo AM, Thavalathil B, Elwyn G, Nemeth Z, Dang S. The Promise of Electronic Health Records to Promote Shared Decision Making: A Narrative Review and a Look Ahead. Med Decis Making 2018 Nov;38(8):1040-1045. [doi: 10.1177/0272989X18796223] [Medline: 30226100]
31. Scalia P, Durand M, Forcino RC, Schubbe D, Barr PJ, O'Brien N, et al. Implementation of the uterine fibroids Option Grid patient decision aids across five organizational settings: a randomized stepped-wedge study protocol. Implement Sci 2019 Sep 02;14(1):88 [FREE Full text] [doi: 10.1186/s13012-019-0933-z] [Medline: 31477140]
32. Yin R. Case Study Research and Applications Design and Methods. Los Angeles: Sage publications; 2017.
33. Rodgers M, Thomas S, Harden M, Parker G, Street A, Eastwood A. Developing a methodological framework for organisational case studies: a rapid review and consensus development process. In: Health Serv Deliv Res. Southampton: NIHR Journals Library; Jan 2016:1-142 URL: https://www.ncbi.nlm.nih.gov/books/NBK333581/pdf/Bookshelf_NBK333581.pdf
34. DynaMed Shared Decisions. EBSCO Health. URL: https://health.ebsco.com/products/option-grid [accessed 2020-06-01]
35. What is SMART? SMART Health IT. URL: https://smarthealthit.org/an-app-platform-for-healthcare/about/ [accessed 2020-06-01]
36. HL7. URL: https://www.hl7.org/ [accessed 2020-06-01]
37. Kamel PI, Nagy PG. Patient-Centered Radiology with FHIR: an Introduction to the Use of FHIR to Offer Radiology a Clinically Integrated Platform. J Digit Imaging 2018 Jun;31(3):327-333 [FREE Full text] [doi: 10.1007/s10278-018-0087-6] [Medline: 29725963]
38. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc 2016 Sep:899-908 [FREE Full text] [doi: 10.1093/jamia/ocv189] [Medline: 26911829]
39. Semenov I, Kopanitsa G, Denisov D, Alexandr Y, Osenev R, Andreychuk Y. Patients Decision Aid System Based on FHIR Profiles. J Med Syst 2018 Jul 31;42(9):166. [doi: 10.1007/s10916-018-1016-4] [Medline: 30066031]
40. Melnick ER, Lopez K, Hess EP, Abujarad F, Brandt CA, Shiffman RN, et al. Back to the Bedside: Developing a Bedside Aid for Concussion and Brain Injury Decisions in the Emergency Department. EGEMS (Wash DC) 2015;3(2):1136. [doi: 10.13063/2327-9214.1136] [Medline: 26290885]
41. Jetelina KK, Woodson TT, Gunn R, Muller B, Clark KD, DeVoe JE, et al. Evaluation of an Electronic Health Record (EHR) Tool for Integrated Behavioral Health in Primary Care. J Am Board Fam Med 2018;31(5):712-723 [FREE Full text] [doi: 10.3122/jabfm.2018.05.180041] [Medline: 30201667]
42. McCrorie C, Benn J, Johnson OA, Scantlebury A. Staff expectations for the implementation of an electronic health record system: a qualitative study using normalisation process theory. BMC Med Inform Decis Mak 2019 Nov 14;19(1):222 [FREE Full text] [doi: 10.1186/s12911-019-0952-3] [Medline: 31727063]
43. Scalia P, Elwyn G, Durand M. "Provoking conversations": case studies of organizations where Option Grid™ decision aids have become 'normalized'. BMC Med Inform Decis Mak 2017 Aug 18;17(1):124 [FREE Full text] [doi: 10.1186/s12911-017-0517-2] [Medline: 28821256]
44. Silvia KA, Sepucha KR. Decision aids in routine practice: lessons from the breast cancer initiative. Health Expect 2006 Sep;9(3):255-264. [doi: 10.1111/j.1369-7625.2006.00393.x] [Medline: 16911140]
45. Uy V, May SG, Tietbohl C, Frosch DL. Barriers and facilitators to routine distribution of patient decision support interventions: a preliminary study in community-based primary care settings. Health Expect 2014 Jul;17(3):353-364 [FREE Full text] [doi: 10.1111/j.1369-7625.2011.00760.x] [Medline: 22212453]

46. Bloomfield RA, Polo-Wood F, Mandel JC, Mandl KD. Opening the Duke electronic health record to apps: Implementing SMART on FHIR. Int J Med Inform 2017 Dec;99:1-10. [doi: 10.1016/j.ijmedinf.2016.12.005] [Medline: 28118917]

Abbreviations

AI: artificial intelligence
EBSCO: Elton B Stephens Company
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
LLC: limited liability company
PCORI: Patient-Centered Outcomes Research Institute
PDA: patient decision aid
PHI: protected health information
SMART: Substitutable Medical Applications and Reusable Technologies
UPFRONT: Uterine Fibroids Options

Edited by R Kukafka; submitted 22.07.20; peer-reviewed by L Lenert, A Hoffman; comments to author 01.09.20; revised version received 20.10.20; accepted 17.02.21; published 03.05.21

Please cite as:
Scalia P, Ahmad F, Schubbe D, Forcino R, Durand MA, Barr PJ, Elwyn G
Integrating Option Grid Patient Decision Aids in the Epic Electronic Health Record: Case Study at 5 Health Systems
J Med Internet Res 2021;23(5):e22766
URL: https://www.jmir.org/2021/5/e22766
doi: 10.2196/22766
PMID:

Original Paper

The Uptake and Use of Telemonitoring in Chronic Care Between 2014 and 2019: Nationwide Survey Among Patients and Health Care Professionals in the Netherlands

Martine W J Huygens[1], MSc, PhD; Helene R Voogdt-Pruis[2], MSc, PhD; Myrah Wouters[3], MA; Maaike M Meurs[1], MSc, PhD; Britt van Lettow[4], MSc, PhD; Conchita Kleijweg[4], MSc; Roland D Friele[1], MSc, PhD

[1]NIVEL Netherlands Institute for Health Services Research, Utrecht, Netherlands

[2]Department of Global Health, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands

[3]National Institute for Public Health and the Environment, Bilthoven, Netherlands

[4]Nictiz, The Hague, Netherlands

Corresponding Author:
Helene R Voogdt-Pruis, MSc, PhD
Department of Global Health
Julius Center for Health Sciences and Primary Care
University Medical Center Utrecht
PO Box 85500
Utrecht, 3508 GA
Netherlands
Phone: 31 0651964815
Email: h.r.pruis@umcutrecht.nl

Abstract

Background: Telemonitoring could offer solutions to the mounting challenges for health care and could improve patient self-management. Studies have addressed the benefits and challenges of telemonitoring for certain patient groups.

Objective: This paper will examine the nationwide uptake of telemonitoring in chronic care in the Netherlands from 2014 to 2019 by means of an annual representative survey among patients and health care professionals.

Methods: Between 2014 and 2019, approximately 2900 patients with chronic diseases, 700 nurses, and 500 general practitioners (GPs) and medical specialists received a questionnaire. About 30 questions addressed topics about the use of eHealth and experiences with it, including data about telemonitoring.

Results: Between 2014 and 2019, the use of telemonitoring remained stable for all groups except medical specialists. In medical specialist departments, the use of telemonitoring increased from 11.2% (18/161) in 2014 to 19.6% (36/184) in 2019 (χ^2_4=12.3; *P*=.02). In 2019, telemonitoring was used by 5.8% (28/485) of people with chronic disease. This was 18.2% (41/225) in GP organizations and 40.4% (44/109), 38.0% (78/205), and 8.9% (29/325) in the organizations of nurses working in primary, secondary, and elderly care, respectively. Up to 10% of the targeted patient group such as diabetics were regarded by health care professionals as suitable for using telemonitoring. The main benefits mentioned by the patients were “comfort” (421/1043, 40.4%) and “living at home for longer/more comfortably” (334/1047, 31.9%). Health care professionals added “improvement of self-management” (63/176, 35.8% to 57/71, 80.3%), “better understanding of the patient’s condition” (47/176, 26.7% to 42/71, 59.2%), “reduction of workload” (53/134, 39.6% of nurses in elderly care), “better tailoring of care plan to the patient’s situation” (95/225, 42.2% of GPs), and “saves time for patients/caregivers” (61/176, 34.7% of medical specialists). Disadvantages mentioned by professionals were that “it takes time to monitor data” (13/130, 10% to 108/225, 48.0%), “it takes time to follow up alerts” (15/130, 11.5% to 117/225, 52.0%), and “it is difficult to estimate which patients can work with telemonitoring” (22/113, 19.5% to 94/225, 41.8%).

Conclusions: The uptake of telemonitoring in Dutch chronic care remained stable during 2014-2019 but increased among medical specialists. According to both patients and professionals, telemonitoring improves the quality of life and quality of care. Skills for suitably including eligible patients and for allocating the tasks of data monitoring and follow-up care within the team would help to further increase the use of telemonitoring.

(J Med Internet Res 2021;23(5):e24908) doi: 10.2196/24908

KEYWORDS

eHealth; telemonitoring; self-management; telemedicine; telehealth

Introduction

Added Value of Telemonitoring

Telemonitoring could broaden access to health care and offer solutions to the mounting challenges for the health care system such as an ageing population, which is creating a demand for long-term care, rising expectations from patients who are better informed about health issues, and the pressure on national health care budgets due to these demands [1-4]. Telemonitoring uses technology such as videoconferencing, email, remote electronic monitoring equipment, social network apps, and internet portals to allow monitoring and self-monitoring of health data by patients and health-related education and long-distance interventions by health care professionals (HCPs) [5-7]. Several studies have addressed the benefits of telemonitoring, for example, better access to health care and the cost-effective delivery of health care. Telemonitoring could reduce face-to-face consultations and clinic visits. In addition, telemonitoring improves the quality of care and clinical outcomes through continuous and reliable monitoring of data, immediate assessment, triage, and interventions. Telemonitoring could also improve patient empowerment, self-management, and compliance [4,8-16].

Implementation of Telemonitoring

Implementation and actual use of telemonitoring in daily practice require well-thought-out action plans, for example, for selecting appropriate interventions or for tailoring the design to the needs of the user group. In addition, telemonitoring should be seamlessly integrated into the health care processes and should avoid disrupting the HCPs' existing workflow. It was also recommended that telemonitoring should be part of "blended care" as both patients and HCPs prefer face-to-face encounters [17,18]. Implementation programs for telemonitoring should tackle barriers to actual use. Some studies found barriers that are related to "users" of telemonitoring such as lack of digital skills, resistance to change, and lack of direct personal benefit. Mentioned examples of barriers that were related to the context were lack of privacy (or fear of it), security, patient safety, a properly working internet connection, proper technological infrastructure, regulations, funding, and task allocation [4,19].

Telemonitoring in the Netherlands

In the Netherlands, an enabling factor related to the context of telemonitoring is the availability of 4G mobile networks and high-speed broadband Internet access, even in rural areas [20]. Currently, 90% of Dutch people use the internet daily. In particular, over the last 5 years, older people, the less well-educated, those born outside the country, and low-income households have caught up [21]. In terms of policy, several national documents, studies, and guidelines for eHealth have been developed during the last 10 years addressing privacy and patient safety in the use of eHealth [22,23]. In 2012, the Dutch National Implementation Agenda for eHealth was launched, followed by the eHealth Governance Covenant 2014 - 2019 [24] and a framework on the use of eHealth by HCPs [23]. As regards funding, telemonitoring is covered by the Dutch health insurance system; in particular, the funding of follow-up consultations improved in 2019. Still, HCPs do not have the funds to upscale telemonitoring [25].

Annual Nationwide Representative Study Among Health Care Professionals

Since 2013, the nationwide uptake of eHealth in general by patients and HCPs has been investigated annually and reported in what is known as the "eHealth-monitor." This investigation was commissioned by the Dutch Ministry of Health, Welfare and Sport. The aim of the annual eHealth-monitor was to investigate the implementation of eHealth and to boost its implementation in subsequent health care policy. Every year, about 30 questions addressed topics on the use of eHealth and experiences with it, including telemonitoring. The findings from the perspectives of nurses, general practitioners (GPs), medical specialists, and patients with a chronic disease increased the understanding of the implementation of telemonitoring and the uptake of telemonitoring in daily practice. To our knowledge, our study is the first scientific paper on the nationwide uptake of telemonitoring for all patient groups in chronic care over a long period of time. We analyzed data on the actual use of telemonitoring by patients and HCPs in daily chronic care. In addition, opinions on and experiences with telemonitoring were analyzed.

Methods

Study Design

Since 2013, data for the eHealth-monitor has been collected annually using various nationwide panels. Written and online questionnaires on eHealth were sent to GPs, medical specialists, nurses (practice nurses and practice assistants working in elderly care, GP care, and hospital care), and people with chronic diseases. All participants were approached in March. Nonresponders initially received 1 written or 2 online reminders. For this study, data from respondents on questions about telemonitoring between 2014 and 2019 were used.

Study Population

People with chronic diseases may be included in the representative National Panel of people with Chronic illness or Disability (NPCD) [26]. Inclusion criteria for the NPCD are age 15 or older, diagnosed with a somatic chronic disease, aware of the diagnosis, having a life expectancy of more than 6 months, mentally capable of participating, and not permanently institutionalized. Every year, 500 new panel members are selected to replace panel members who have withdrawn or who have participated for 4 years. The questions on telemonitoring were posed in 2015, 2017, and 2019.

Nurses are participants in a representative nursing staff panel. The nursing staff panel consists of a nationwide group of nursing staff members (nurses, caregivers, and practice assistants) in various health care settings who deliver direct patient care. The recruitment of members of the nursing staff panel takes place through a random sample of two pension funds. Together, these pension funds register all employees in the Dutch health care sector. Nursing staff were asked to participate in health care research for various purposes. People who agreed and who delivered direct nursing care to patients could join the nursing staff panel. The questions on telemonitoring were asked annually, from 2014 until 2019. For this study, the data of nurses working in primary care, secondary care, and elderly care were used. For 2014 until 2016, the data of nurses working in the curative sectors (practice nurses and nurses working in hospital care) were taken together. From 2017 onwards, these sectors were split into two samples.

General practitioners and medical specialists are participants in a representative doctor's panel. Included are all registered GPs and medical specialists of the Royal Dutch Medical Association. Inclusion criteria for participating in the eHealth-monitor were practicing in the past year and being involved in the diagnosis or treatment of patients. From these doctors, certain specializations were excluded: public and occupational health, forensic medicine, addiction medicine, and psychiatry. The questions on telemonitoring were asked annually, from 2014 until 2019.

Questionnaires

All participants were asked about their use of telemonitoring and experience with its advantages and disadvantages in the previous 12 months. In addition, HCPs were asked for which portion of patients telemonitoring was used and for which patient groups relevant (Textbox 1).

Textbox 1. Questionnaires.

Telemonitoring: Remote monitoring of a patient, in which they measure their own health values (for example, blood pressure, blood sugar level) using a meter, sensor, or other device in the home situation and in which they could also respond to some questions. The HCP receives these data digitally.

People with chronic diseases who measure health values themselves

- Which of the following statements apply to you? (multiple answers possible)
- I electronically submit my self-measured health values to my health care provider (eg, by email or automatically via computer or mobile app) (2015, 2017, 2019)
- My health care provider can see my health data on a website or in a mobile app (2015, 2017, 2019)
- My health care provider looks at my self-measured health data before or during a consultation and discusses it with me (2015, 2017, 2019)
- My health care provider keeps an eye on my self-measured health data remotely and contacts me if anything is wrong (2015, 2017, 2019)
- Can you say how desirable or necessary telemonitoring is for you? (2017)
- Could you please answer the following statements? (2019) I notice or think that telemonitoring...
  - ...makes it easier for me to live at home longer and/or more easily
  - ...takes a lot of effort for me
  - ...improves my care
  - ...makes me very tense

Nurses

- Has telemonitoring been used in your organization in the past year? (2014-2019)
- If so, what proportion of your clients/patients use telemonitoring (estimated)? (2016, 2017, 2019)
- If so, what proportion of your clients do you think telemonitoring makes sense for (estimated)? (2019)
- What advantages and disadvantages do you experience with telemonitoring or expect from it? (multiple answers possible) (2019)

General practitioners

- Could you state whether telemonitoring is applied to the following patient groups (in your practice)? If telemonitoring is not applied, could you please state whether there are plans to start within a year and whether you would like to do so? (2014-2017)
- In your practice, some of your patients with diabetes are currently being monitored by telemonitoring. Could you please estimate the proportion of your diabetes patients for whom you think telemonitoring is sensible? What proportion of your diabetes patients do use it? (2015-2017)
- Is telemonitoring of patients relevant to you or your practice, and is it used within your practice for some or all of the patients? (2019)
- Could you estimate the size of the group of patients for whom you think telemonitoring is sensible and for whom you or other HCPs actually use it in your practice? (2019)
- What advantages and disadvantages do you experience with telemonitoring or expect from it? (multiple answers possible) (2019)

Medical specialists

- Is telemonitoring of patients relevant to your medical specialty, and is it applied by your department for some or all patients? (2014-2017, 2019)
- For which group of your patients is telemonitoring relevant? (2015-2017)
- Could you estimate the size of the group of patients for whom you think telemonitoring is sensible and for whom you or other HCPs actually use it at your department? (2015-2017, 2019)
- What advantages and disadvantages do you experience with telemonitoring or expect from it? (multiple answers possible) (2019)

Data Analysis

Descriptive analyses were conducted to study the use and experiences of telemonitoring. Data from the questionnaires among people with a chronic disease and nursing staff were analyzed using Stata, version 15.0 (StataCorp). The results from the questionnaires among GPs and medical specialists were analyzed using SPSS, version 25.0 (IBM Corp).

For questions asked to people with chronic diseases, the descriptive analyses were weighted for age and gender in such a way that it resembled the distribution of age and gender within the Dutch population from age 18 years, based on data from Statistics Netherlands. We applied a weighting factor ranging from 0.65 to 2.28. The samples for nurses and GPs are fairly representative of the Dutch population of nurses and GPs regarding gender, but the response is not representative for age: GPs younger than 35 years and GPs and nurses 50 years and

older responded more often. Nevertheless, we did not use a weight factor to correct for this because applying the weight factor did not affect the results. For questions asked to medical specialists, the descriptive analyses were weighted for type of specialty. We applied a weighting factor ranging from 0.5 to 1.7.

Results

Participants

Over the years, data were used from 485-633 people with a chronic disease, 322-607 nurses working in elderly care, 220-367 nurses working in the curative disciplines (primary and secondary care), 225-396 GPs, and 184-386 medical specialists who answered questions about telemonitoring (Table 1). The mean age of people with chronic disease was 64.5 to 66.4 years. The mean age of nurses varied from 46.7 to 52.2 years. The mean ages of GPs and medical specialists ranged from 50.0 to 52.6 years and 49.4 to 53.9 years, respectively. Approximately half of the people with chronic disease and the doctors were men; among nurses, only 3.5% (21/607) to 16.1% (25/155) were men.

Table 1. Study population per year.[a]

Characteristic	2014	2015	2016	2017	2018	2019
People with chronic disease						
Responses, n	__[b]	1448	—	1357	—	1292
Telemonitoring responses, n[c]	—	604	—	633	—	485
Male, n (%)	—	315 (52.2)	—	304 (48.0)	—	234 (48.2)
Age (years), mean (SD)	—	64.9 (12.7)	—	66.4 (12.1)	—	64.5 (11.9)
Low education, n (%)	—	133 (22.0)	—	156 (24.6)	—	95 (19.6)
Nurses in elderly care						
Telemonitoring responses, n	408	607	433	341	322	325
Male, n (%)	18 (4.4)	21 (3.5)	20 (4.6)	15 (4.4)	29 (9.0)	19 (5.8)
Age (years), mean (SD)	50.8 (8.9)	48.9 (10.8)	49.7 (10.3)	52.2 (9.5)	50.1 (10.8)	50.9 (11.1)
Nurses in curative sectors						
Telemonitoring responses, n	262	316	220	—	—	—
Male, n (%)	26 (9.9)	36 (11.4)	25 (11.4)	—	—	—
Age (years), mean (SD)	48.8 (9.6)	46.7 (11.4)	48.4 (10.7)	—	—	—
Practice nurses						
Telemonitoring responses, n	—	—	—	212	124	109
Male, n (%)	—	—	—	12 (5.7)	8 (6.5)	8 (7.3)
Age (years), mean (SD)	—	—	—	50.8 (8.6)	50.4 (9.0)	52.0 (9.1)
Nurses in hospital care						
Telemonitoring responses, n	—	—	—	155	211	205
Male, n (%)	—	—	—	25 (16.1)	29 (13.7)	27 (13.2)
Age (years), mean (SD)	—	—	—	49.4 (11.0)	50.0 (11.2)	48.2 (12.4)
General practitioners						
Telemonitoring responses, n	—	396	316	290	—	225
Male, n (%)	—	194 (49.0)	162 (51.3)	133 (45.9)	—	106 (47.1)
Age (years), mean (SD)	—	50.0 (10.2)	51.1 (9.9)	51.5 (9.7)	—	52.6 (9.1)
Medical specialists						
Telemonitoring responses, n[d]	—	386	274	253	—	184
Male, n (%)	—	203 (52.6)	151 (55.1)	153 (60.5)	—	118 (64.1)
Age (years), mean (SD)	—	49.8 (11.3)	49.4 (11.1)	51.9 (11.3)	—	53.9 (9.7)

[a]Among general practitioners and medical specialists, the questions in 2015, 2016, and 2017 were different from those in 2019.

[b]Not available.

[c]Question was only asked to persons who measured health outcomes by themselves.

[d]Weighted by type of specialty.

Actual Use of Telemonitoring

In 2019, 5.8% (28/485) of people with a chronic disease stated that their HCP monitors the health values remotely and contacts them if anything looks wrong (Table 2). In 2015, this figure was 3.3% (20/604). Most of them had been diagnosed with cardiovascular disease or diabetes. In 2019, 38.0% (78/205) of nurses working in secondary care and 40.4% (44/109) in primary care stated that their organization uses telemonitoring. In 2014, this percentage was 34.0% (89/262) for nurses working in the curative disciplines (both hospital and general practice nurses). Among nurses working in elderly care this was 8.9% (29/325) in 2019 and 11.3% (46/408) in 2014. In addition, 18.2% (41/225)of GPs and 19.6% (36/184)of medical specialists stated that telemonitoring was used in their organizations in 2019. These percentages were 17.0% and 11.2%, respectively, in 2014. Among medical specialists, this number has grown significantly over the years (χ^2_4=12.3, P=.02). Up to 10% of the targeted

patient group, including people with diabetes, is reckoned to be suitable for using telemonitoring.

Table 2. Proportions of patients and health care professionals (HCPs) using telemonitoring, from 2014 to 2019.

Population using telemonitoring	2014	2015	2016	2017	2018	2019
Patients with a chronic disease, n (%)[a]	—[b]	20 (3.3)	—	37 (5.8)	—	28 (5.8)
Elderly care						
Nurses, n (%)[c]	46 (11.3)	70 (11.5)	—	—	—	29 (8.9)
Patients using telemonitoring, n (mode %)[d]	—	—	27 (1-10)	25 (1-10)	—	17 (1-10)
Primary/secondary care						
Nurses, n (%)	89 (34.0)	126 (39.9)	84 (38.2)	—	—	—
Patients using telemonitoring, n (mode %)	—	—	18 (1-10)	—	—	—
Primary care						
Nurses, n (%)	—	—	—	85 (40.1)	58 (46.8)	44 (40.4)
Patients using telemonitoring, n (mode %)	—	—	—	61 (1-10)	—	25 (1-10)
Secondary care						
Nurses, n (%)	—	—	—	50 (32.3)	80 (37.9)	78 (38.0)
Patients using telemonitoring, n (mode %)	—	—	—	14 (1-10)	—	22 (1-10)
General practitioners, n (%)[c]	29 (17.0)	49 (12.4)	41 (13.0)	26 (9.0)	—	41 (18.2)
Group using telemonitoring, n (mode group)	25 (Diabetes)	43 (Diabetes)	37 (Diabetes)	21 (Diabetes)	—	—
Relevant for diabetes, n (mode %)[d]	—	26 (1-10)	20 (1-10)	10 (1-10)	—	—
Medical specialists, n (%)[c]	18 (11.2)	41 (10.6)	26 (9.5)	29 (11.5)	—	36 (19.6)
Relevant for patients, n (mode %)[d]	—	24 (1-10)	19 (1-10)	—	—	—
Relevant for patient group, n (mode group)[e]	—	13 (Diabetes)	9 (Diabetes)	11 (Diabetes)	—	—

[a]The proportion of responding patients stating they are monitored on self-reported health measures by HCP (remotely).

[b]Not available.

[c]The proportion of responding HCPs stating that telemonitoring is relevant or is used at their department, practice, or organization. Difference between years (χ^2_4=12.3; *P*=.02).

[d]Proportion of patients using telemonitoring according to the HCPs; possible answers: none, up to 10%, up to 20%, up to 50%, up to 100%, I don't know.

[e]The most often mentioned were these groups: diabetes, heart failure, chronic obstructive pulmonary disease, and asthma.

Advantages and Disadvantages of Telemonitoring

Of patients with a chronic disease, 40.4% (421/1043) agreed or totally agreed with the statement "telemonitoring improves my comfort" (353/1043, 33.8% answered "don't know"). In addition, 31.9% (334/1047) agreed or totally agreed with "telemonitoring lets me stay at home longer and/or live more comfortably" (405/1050, 38.7% answered "don't know") (Table 3).

On the other hand, 7.7% agreed or totally agreed with "telemonitoring takes me a lot of effort" (36.4% answered "don't know"). Among HCPs, the most widely experienced or expected advantage of telemonitoring mentioned was "telemonitoring improves patients' self-management" (57/71, 80.3% of nurses working in primary care; 71/124, 57.3% of nurses working in secondary care; 84/134, 62.7% of nurses working in elderly care; 135/225, 60% of GPs; 63/176, 35.8% of medical specialists) (Table 4). In addition, 59.2% (42/71) of nurses working in primary care and 46.8% (58/124) of nurses working in secondary care stated that their experience or expectation of telemonitoring was that they would get a better understanding of the health condition of the patient. Of nurses working in elderly care, 39.6% (53/134) experienced or expected telemonitoring to reduce the workload. Moreover, 42.2% (95/225) of GPs said that telemonitoring let them tailor the care plan to the situation of their patients better. Of medical specialists, 34.7% (61/176) expected or had observed that telemonitoring gave them spare time for patients and caregivers/relatives.

Table 3. Opinions of telemonitoring from patients with chronic diseases (n=1023-1050), 2019.

I notice or think that telemonitoring:	Agreement with statement, n (%)					
	Totally disagree	Disagree	Neither agree nor disagree	Agree	Totally agree	I don't know
Lets me live at home longer and/or more easily (n=1047)	47 (4.5)	62 (5.9)	199 (19.0)	235 (22.4)	99 (9.5)	405 (38.7)
Improves my comfort (n=1043)	42 (4.0)	50 (4.8)	177 (17.0)	331 (31.7)	90 (8.6)	353 (33.8)
Takes me a lot of effort (n=1023)	83 (8.1)	249 (24.3)	246 (24.0)	61 (6.0)	12 (1.2)	372 (36.4)
Improves my care (n=1035)	41 (4.0)	66 (6.4)	234 (22.6)	241 (23.3)	82 (7.9)	371 (35.8)
Makes me very tense (n=1027)	104 (10.1)	220 (21.4)	238 (23.2)	77 (7.5)	17 (1.7)	371 (36.1)

Table 4. Experienced or expected advantages and disadvantages of telemonitoring assessed by health care providers, 2019.

Characteristic	Nurses, primary care	Nurses, secondary care	Nurses, elderly care	General practitioners	Medical specialists
Advantages respondents, n	71	124	134	225	176
Advantages, n (%)[a]					
It reduces the workload	10 (14.1)	23 (18.5)	53 (39.6)	16 (7.1)	14 (8.0)
It reduces the workload of my assistants	__[b]	—	—	56 (24.9)	18 (10.2)
It saves time for patients and or caregivers/relatives	32 (45.1)	58 (46.8)	45 (33.6)	86 (38.2)	61 (34.7)
It improves the quality of care in my organization	34 (47.9)	50 (40.3)	38 (28.4)	77 (34.2)	44 (25.0)
It improves the self-management of the patient	57 (80.3)	71 (57.3)	84 (62.7)	135 (60.0)	63 (35.8)
I have a better understanding of my clients' health condition	42 (59.2)	58 (46.8)	43 (32.1)	77 (34.2)	47 (26.7)
It lets me tailor the care plan better to my patients' situation	38 (53.5)	47 (37.9)	32 (23.9)	95 (42.2)	45 (25.6)
It lets patients ask for help in time	20 (28.2)	52 (41.9)	38 (28.4)	67 (29.8)	32 (18.2)
Other advantages	—	7 (5.6)	3 (2.2)	9 (4.0)	5 (2.8)
I do not expect or experience any advantages	—	9 (7.3)	23 (17.2)	32 (14.2)	48 (27.3)
Disadvantages respondents, n	68	113	130	225	181
Disadvantages, n (%)[a]					
It takes me a lot of time to monitor/check health values	32 (47.1)	21 (18.6)	13 (10.0)	108 (48.0)	56 (30.9)
It takes me a lot of time to follow up notifications	29 (42.6)	19 (16.8)	15 (11.5)	117 (52.0)	51 (28.2)
It takes a lot of time for my assistants	—	—	—	99 (44.0)	45 (24.9)
It ensures that patients and/or relatives contact me more often	20 (29.4)	22 (19.5)	21 (16.2)	83 (36.9)	40 (22.1)
It worries patients and/or relatives	10 (14.7)	20 (17.7)	23 (17.7)	81 (36.0)	33 (18.2)
I find it difficult to work with it	—	—	—	16 (7.1)	7 (3.9)
The system provides unreliable data	—	—	—	52 (23.1)	33 (18.2)
The application is not secure	—	—	—	20 (8.9)	7 (3.9)
I find it difficult to estimate which patients can work with it	15 (22.1)	22 (19.5)	37 (28.5)	94 (41.8)	52 (28.7)
Other disadvantages	8 (11.8)	7 (6.2)	8 (6.2)	18 (8.0)	20 (11.0)
I do not expect or experience any disadvantages	13 (19.1)	49 (43.4)	56 (43.1)	13 (5.8)	33 (18.2)

[a]Multiple answers possible.

[b]Not available.

A total of 43.4% (49/113) of nurses working in secondary care and 43.1% (56/130) of nurses working in elderly care did not expect or experience any disadvantages of telemonitoring. In contrast, 47.1% (32/68) of nurses working in primary care

expected or noted that telemonitoring takes a lot of time to monitor and check health values. Of GPs and medical specialists, 48.0% (108/225) and 30.9% (56/181), respectively, expected or experienced this. In addition, 42.6% (29/68) of nurses working in primary care, 52.0% (117/225) of GPs, and 28.2% (51/181) of medical specialists expected or experienced that using telemonitoring would take up a lot of time following up on alerts. Moreover, 19.5% (22/113) of nurses working in secondary care, 28.5% (37/130) of the nurses working in elderly care, and 28.7% (52/181) of medical specialists expected or experienced difficulties in estimating which patients could handle telemonitoring.

Discussion

Principal Findings

Our study adds new insights to current scientific studies of telemonitoring, as it investigated the actual nationwide uptake of telemonitoring for all patient groups in chronic care over a long period of time (before the COVID-19 pandemic). Findings from the perspectives of nurses, GPs, medical specialists, and patients with chronic diseases can assist the implementation of telemonitoring and the uptake of telemonitoring in daily practice. The current COVID-19 pandemic has called for rapid implementation of telemonitoring for acute and subacute diseases in 2020. Future editions of the Dutch eHealth-monitor might present the impact of COVID-19 on a rising uptake of telemonitoring [27] and on new opinions and experiences with telemonitoring. Strengths of our study include the large sample size, the external validity and reliability of the data, and the representativeness of the various groups of participants (people with chronic diseases, nurses, GPs, and medical specialists). Nevertheless, there are also some limitations. Due to the specific factors in the Netherlands that boosted the implementation of telemonitoring, our study results can only partly be extrapolated to other countries. In addition, the use of telemonitoring was only investigated by asking potential users in health care; we did not investigate data from other resources such as health care insurers or telehealth companies. As well, our quantitative approach is best suited to answering "what," "when," and "who" questions and less well-suited to "how" and "why" questions. The opinions and experiences with telemonitoring that we investigated therefore do not fully explain the factors concerning the implementation and uptake of telemonitoring. Even so, our results concerning the experiences of HCPs are underlined by qualitative studies [28-30]. Other benefits and barriers found are "an increased feeling of safety" and "insufficient familiarity with the technology" [28-30]. In addition, HCPs need to add telemonitoring to the health care process, with a precise description of the target group, task allocation for data monitoring, and support for patients from within the team [31,32].

Conclusion

The uptake of telemonitoring in Dutch chronic care remained stable during 2014-2019 but increased among medical specialists. According to both patients and professionals, telemonitoring improves the quality of life and quality of care. Skills for appropriately including eligible patients and allocating the tasks of data monitoring and follow-up care within the team would help to further increase the use of telemonitoring.

Acknowledgments

The project was funded by the Dutch Ministry of Health, Welfare and Sport. The Dutch Ministry of Health, Welfare and Sport had no role in the collection, analysis, or interpretation of data or in writing the manuscript. The authors would like to acknowledge all other people who were involved in data collection during 2013-2019; in particular, Johan Krijgsman, Ilse Swinkels, Judith de Jong, Janke de Groot, Anne Brabers, Lies van Gennip, Karlijn de Bruin, and Annemarie Lamain. In addition, the authors would like to thank the members of the advisory committee in 2019: Guus Schrijvers, Christine Aberson-Blok, Arina Burghouts, Jeroen Crasborn, Henny Mulders, Henk Herman Nap, Sjaak Nouwt, Renée Verwey, Karin Zwager-Ankone, and Nienke Zwennes.

Authors' Contributions

MWJH and HRVP were the main contributors in data analysis and in writing the manuscript. MW, MWJH, BvL, CK, and RDF were involved in developing the questionnaires. All authors were involved in the process of data management. MWJH, HRVP, MMM, MW, and BvL were involved in the data analysis. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005 Feb 24;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
2. Epstein RM, Fiscella K, Lesser CS, Stange KC. Why the nation needs a policy push on patient-centered health care. Health Aff (Millwood) 2010 Aug;29(8):1489-1495. [doi: 10.1377/hlthaff.2009.0888] [Medline: 20679652]
3. European Commission, Liikanen E. The Importance of eHealth in Europe. 2003. URL: https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_03_608 [accessed 2020-01-03]
4. World Health Organization. From innovation to implementation: eHealth in the WHO European Region. 2016. URL: https://www.euro.who.int/__data/assets/pdf_file/0012/302331/From-Innovation-to-Implementation-eHealth-Report-EU.pdf [accessed 2020-01-07]

5. Telehealth programs. 2018. URL: https://www.hrsa.gov/rural-health/telehealth/index.html [accessed 2019-11-01]
6. Larsen SB, Sørensen NS, Petersen MG, Kjeldsen GF. Towards a shared service centre for telemedicine: Telemedicine in Denmark, and a possible way forward. Health Informatics J 2016 Dec;22(4):815-827 [FREE Full text] [doi: 10.1177/1460458215592042] [Medline: 26261216]
7. Anker SD, Koehler F, Abraham WT. Telemedicine and remote management of patients with heart failure. Lancet 2011 Aug 20;378(9792):731-739. [doi: 10.1016/S0140-6736(11)61229-4] [Medline: 21856487]
8. Lopez-Villegas A, Catalan-Matamoros D, Robles-Musso E, Bautista-Mesa R, Peiro S. Cost-utility analysis on telemonitoring of users with pacemakers: The PONIENTE study. J Telemed Telecare 2019 May;25(4):204-212. [doi: 10.1177/1357633X18767184] [Medline: 29621908]
9. Boyne JJ, Vrijhoef HJ, Spreeuwenberg M, De Weerd G, Kragten J, Gorgels AP, TEHAF investigators. Effects of tailored telemonitoring on heart failure patients' knowledge, self-care, self-efficacy and adherence: a randomized controlled trial. Eur J Cardiovasc Nurs 2014 Jun;13(3):243-252. [doi: 10.1177/1474515113487464] [Medline: 23630403]
10. Sul A, Lyu D, Park D. Effectiveness of telemonitoring versus usual care for chronic obstructive pulmonary disease: A systematic review and meta-analysis. J Telemed Telecare 2020 May;26(4):189-199. [doi: 10.1177/1357633X18811757] [Medline: 30541375]
11. Qian W, Lam TT, Lam HHW, Li C, Cheung YT. Telehealth Interventions for Improving Self-Management in Patients With Hemophilia: Scoping Review of Clinical Studies. J Med Internet Res 2019 Jul 10;21(7):e12340 [FREE Full text] [doi: 10.2196/12340] [Medline: 31293241]
12. Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff (Millwood) 2014 Feb;33(2):194-199. [doi: 10.1377/hlthaff.2013.0992] [Medline: 24493760]
13. 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 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
14. Nissen L, Lindhardt T. A qualitative study of COPD-patients' experience of a telemedicine intervention. Int J Med Inform 2017 Nov;107:11-17. [doi: 10.1016/j.ijmedinf.2017.08.004] [Medline: 29029687]
15. Mathavakkannan S, Bowser M, Doyle S, Hoare A, Rehsi G. Promoting patient empowerment and sustainability in kidney care using telemedicine. Int J Integr Care 2014 Nov 01;14(8):2. [doi: 10.5334/ijic.1816]
16. van der Burg JM, Aziz NA, Kaptein MC, Breteler MJ, Janssen JH, van Vliet L, et al. Long-term effects of telemonitoring on healthcare usage in patients with heart failure or COPD. Clinical eHealth 2020;3:40-48. [doi: 10.1016/j.ceh.2020.05.001]
17. Ziefle M, Klack L, Wilkowska W, Holzinger A. Acceptance of telemedical treatments – A medical professional point of view. 2013 Presented at: International conference on human interface and the management of information; 2013; Berlin, Germany p. 21-26. [doi: 10.1007/978-3-642-39215-3_39]
18. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for Patient Engagement: A Systematic Review. Front Psychol 2015;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
19. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
20. Eurostat. Digital economy and society statistics - households and individuals. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/Digital_economy_and_society_statistics_-_households_and_individuals [accessed 2020-01-07]
21. Statistics Netherlands (CBS), Eurostat. ICT use by households and individuals. URL: https://opendata.cbs.nl/statline/#/CBS/en/?fromstatweb [accessed 2019-11-01]
22. National Institute for Public Health and Environment (RIVM), Ossebaard HC, de Bruijn ACP, van Gemert-Pijnen JEWC, Geertsma RE. Risk related to the use of ehealth technologies; an explorative study. 2012. URL: https://www.rivm.nl/bibliotheek/rapporten/360127001.pdf [accessed 2020-01-07]
23. Inspectorate on health and youth care (IGJ). Toetsingskader IGJ; inzet van e-health door zorgaanbieders. Report. 2018. URL: https://www.igj.nl/binaries/igj/documenten/toetsingskaders/2019/10/18/toetsingskader-inzet-van-e-health-door-zorgaanbieders/Toetsingskader_IGJ_Inzet+e-health+door+zorgaanbieders.pdf [accessed 2020-01-07]
24. Nivel, Kroneman M, Boerma W, van den Berg M, Groenewegen P, de Jong J, et al. Health systems in transition. 2016. URL: https://www.euro.who.int/__data/assets/pdf_file/0016/314404/HIT_Netherlands.pdf [accessed 2020-01-07]
25. NIVEL, Meurs M, Huygens M, Keuper J, de Groot J, Friele R. Slimme zorg thuis bij COPD en hartfalen Ervaringen en ambities van vijf voorloper initiatieven In Dutch. 2019. URL: https://nivel.nl/sites/default/files/bestanden/1003591.pdf [accessed 2020-01-07]
26. Rijken M, Bensing J. The National Panel of Patients with Chronic Diseases: An infrastructure for patient-centered research. TSG Tijdschrift Voor Gezondheidswetenschappen 2000;78:5.
27. Lai SHS, Tang CQY. Telemedicine and COVID-19: beyond just virtual consultations - the Singapore experience. Bone Jt Open 2020 Jun;1(6):203-204 [FREE Full text] [doi: 10.1302/2633-1462.16.BJO-2020-0042.R1] [Medline: 33225290]

28. Van der Zweth J, Askari M, Brinkhuis M, van Lettow B. The Added Value of Telehomecare for Heart Patients, Healthcare Professionals, and Healthcare Organizations. 2017. URL: https://dspace.library.uu.nl/bitstream/handle/1874/354694/van%20der%20Zweth%20-%205675383%20-%20Thesis.pdf [accessed 2020-01-07]
29. Morton K, Dennison L, May C, Murray E, Little P, McManus RJ, et al. Using digital interventions for self-management of chronic physical health conditions: A meta-ethnography review of published studies. Patient Educ Couns 2017 Apr;100(4):616-635 [FREE Full text] [doi: 10.1016/j.pec.2016.10.019] [Medline: 28029572]
30. Slevin P, Kessie T, Cullen J, Butler MW, Donnelly SC, Caulfield B. A qualitative study of chronic obstructive pulmonary disease patient perceptions of the barriers and facilitators to adopting digital health technology. Digit Health 2019;5:2055207619871729 [FREE Full text] [doi: 10.1177/2055207619871729] [Medline: 31489206]
31. Hallensleben C, van Luenen S, Rolink E, Ossebaard HC, Chavannes NH. eHealth for people with COPD in the Netherlands: a scoping review. Int J Chron Obstruct Pulmon Dis 2019;14:1681-1690 [FREE Full text] [doi: 10.2147/COPD.S207187] [Medline: 31440044]
32. Parker S, Prince A, Thomas L, Song H, Milosevic D, Harris MF, IMPACT Study Group. Electronic, mobile and telehealth tools for vulnerable patients with chronic disease: a systematic review and realist synthesis. BMJ Open 2018 Aug 29;8(8):e019192 [FREE Full text] [doi: 10.1136/bmjopen-2017-019192] [Medline: 30158214]

Abbreviations

GP: general practitioner
HCP: health care professional
NPCD: National Panel of people with Chronic illness or Disability

Edited by R Kukafka; submitted 09.10.20; peer-reviewed by K Fitzner, O Blanson Henkemans; comments to author 23.11.20; revised version received 27.11.20; accepted 21.03.21; published 03.05.21

Please cite as:
Huygens MWJ, Voogdt-Pruis HR, Wouters M, Meurs MM, van Lettow B, Kleijweg C, Friele RD
The Uptake and Use of Telemonitoring in Chronic Care Between 2014 and 2019: Nationwide Survey Among Patients and Health Care Professionals in the Netherlands
J Med Internet Res 2021;23(5):e24908
URL: https://www.jmir.org/2021/5/e24908
doi: 10.2196/24908
PMID:

Original Paper

Using a Secure, Continually Updating, Web Source Processing Pipeline to Support the Real-Time Data Synthesis and Analysis of Scientific Literature: Development and Validation Study

Uddhav Vaghela[1*], MBBS, BSc; Simon Rabinowicz[1*], MBBS, BSc; Paris Bratsos[1], MBBS, BSc; Guy Martin[1], PhD; Epameinondas Fritzilas[2], PhD; Sheraz Markar[1], PhD; Sanjay Purkayastha[1], PhD; Karl Stringer[3]; Harshdeep Singh[4], MSc; Charlie Llewellyn[2]; Debabrata Dutta[4], PhD; Jonathan M Clarke[1], PhD; Matthew Howard[2], PhD; PanSurg REDASA Curators[1,5*]; Ovidiu Serban[6], PhD; James Kinross[1], FRCS, PhD

[1]PanSurg Collaborative, Department of Surgery and Cancer, Imperial College London, London, United Kingdom

[2]Amazon Web Services UK Limited, London, United Kingdom

[3]MirrorWeb Limited, Manchester, United Kingdom

[4]Cloudwick Technologies, Newark, CA, United States

[5]see Acknowledgments, London, United Kingdom

[6]Data Science Institute, Imperial College London, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Ovidiu Serban, PhD
Data Science Institute
Imperial College London
William Penney Laboratory
South Kensington Campus
London
United Kingdom
Email: o.serban@imperial.ac.uk

Abstract

Background: The scale and quality of the global scientific response to the COVID-19 pandemic have unquestionably saved lives. However, the COVID-19 pandemic has also triggered an unprecedented “infodemic”; the velocity and volume of data production have overwhelmed many key stakeholders such as clinicians and policy makers, as they have been unable to process structured and unstructured data for evidence-based decision making. Solutions that aim to alleviate this data synthesis–related challenge are unable to capture heterogeneous web data in real time for the production of concomitant answers and are not based on the high-quality information in responses to a free-text query.

Objective: The main objective of this project is to build a generic, real-time, continuously updating curation platform that can support the data synthesis and analysis of a scientific literature framework. Our secondary objective is to validate this platform and the curation methodology for COVID-19–related medical literature by expanding the COVID-19 Open Research Dataset via the addition of new, unstructured data.

Methods: To create an infrastructure that addresses our objectives, the PanSurg Collaborative at Imperial College London has developed a unique data pipeline based on a web crawler extraction methodology. This data pipeline uses a novel curation methodology that adopts a human-in-the-loop approach for the characterization of quality, relevance, and key evidence across a range of scientific literature sources.

Results: REDASA (Realtime Data Synthesis and Analysis) is now one of the world’s largest and most up-to-date sources of COVID-19–related evidence; it consists of 104,000 documents. By capturing curators’ critical appraisal methodologies through the discrete labeling and rating of information, REDASA rapidly developed a foundational, pooled, data science data set of over 1400 articles in under 2 weeks. These articles provide COVID-19–related information and represent around 10% of all papers about COVID-19.

Conclusions: This data set can act as ground truth for the future implementation of a live, automated systematic review. The three benefits of REDASA's design are as follows: (1) it adopts a user-friendly, human-in-the-loop methodology by embedding an efficient, user-friendly curation platform into a natural language processing search engine; (2) it provides a curated data set in the JavaScript Object Notation format for experienced academic reviewers' critical appraisal choices and decision-making methodologies; and (3) due to the wide scope and depth of its web crawling method, REDASA has already captured one of the world's largest COVID-19–related data corpora for searches and curation.

(J Med Internet Res 2021;23(5):e25714) doi: 10.2196/25714

KEYWORDS

structured data synthesis; data science; critical analysis; web crawl data; pipeline; database; literature; research; COVID-19; infodemic; decision making; data; data synthesis; misinformation; infrastructure; methodology

Introduction

Between December 31, 2019, and August 3, 2020, 37,362 papers related to COVID-19 were published on PubMed [1], with Dimensions reporting over 100,743 publications, 1503 policy documents, and 1097 data sets [2]. The speed and scale of the production of published data on COVID-19 across both peer- and nonpeer-reviewed literature presents considerable challenges for stakeholders (eg, policy makers, clinicians, and patients) who must make subjective quality judgements on new data and rapidly synthesize information in order to make optimal, evidence-based decisions. Traditional approaches to data synthesis are unable to keep pace with the rapidly changing information landscape. For example, in the United Kingdom, the National Institute for Health and Care Excellence was unable to publish their initial therapeutic guidance on managing COVID-19 until March 20, 2019 [3]. Ultimately, they modified their methodology for publishing rapid guidance materials on COVID-19 [4]. Moreover, there have been concerns regarding data credibility and the political misuse of information, resulting in the World Health Organization announcing its campaign for discouraging the spread of misinformation [5]. The COVID-19 pandemic highlights the urgent need to prospectively capture, structure, and interpret expansive and complex data sets in real time to support the rapid development of clinical guidance and, most critically, ensure that various key stakeholders can make the best possible evidence-based decisions during an "infodemic."

Previous strategies have attempted to address these challenges by using the concepts of live systematic reviews, which involve the continuous, structured analysis of data that target specific clinical questions [6,7] as well as the clear presentation of such data [8]. However, despite the progress in this field, major obstacles remain in establishing automated data mining frequency, depth, and robustness. Moreover, major barriers exist in the development and validation of machine learning methodologies for the autonomous analysis of heterogeneous clinical data sets.

This paper outlines the methodology of REDASA (Realtime Data Synthesis and Analysis)—a novel, prospective clinical information platform that was developed during the COVID-19 pandemic. It was designed for use across a wide range of data-rich subject areas while keeping application and impact in mind. Our objective was to continuously capture and synthesize both academic and relevant grey literature (eg, news websites, policy documentation, and social media posts) on COVID-19 and to develop a validated data curation approach that could supplement machine learning methodologies and be used as the basis for conducting live systematic reviews.

Methods

Components of REDASA

REDASA was built and deployed on the Amazon Web Service (AWS) cloud. Cloud computing is the on-demand delivery of compute power, database storage, applications, and other information technology resources through a cloud service platform on the internet. A cloud services platform, such as AWS, owns and maintains the network-connected hardware required for these application services. By using the AWS cloud, REDASA components were rapidly designed. REDASA components were integrated and deployed by using AWS tools and the solutions developed by AWS Partners, MirrorWeb, and Cloudwick. These components were comprised of a real-time data extraction pipeline that was implemented by using MirrorWeb's digital archiving technology, a data lake storage repository and workflow orchestration platform (Amorphic) that was developed by Cloudwick, a natural language search engine that was implemented by using Amazon Kendra, and a document curation pathway that was implemented by using Amazon SageMaker Ground Truth.

Real-Time Data Extraction Pipeline

MirrorWeb was used to conduct an exploratory review of the target websites via manual and automatic content detection for informing crawl scoping decisions. Exploratory reviews involve the domain composition analysis of initial web estate archives, which can be produced via multiple methods, including basic link harvesting, Domain Name System lookups, the gathering of URL lists at crawl time (to identify content delivery networks and perform manual verifications), and the inspection of websites. This ensures that (1) the relevant areas of websites are identified and followed by the archive tools and (2) content that can be confidently omitted is avoided. With the adoption of machine learning algorithms, this process can be further assisted by technology.

Scoping decisions can encompass a range of factors. For instance, scoping decisions can ensure that the website crawl stays within the target domain. Further, they can refine the website crawl to only include relevant URL patterns within the domain. For example, if there is a domain.com/coronavirus/

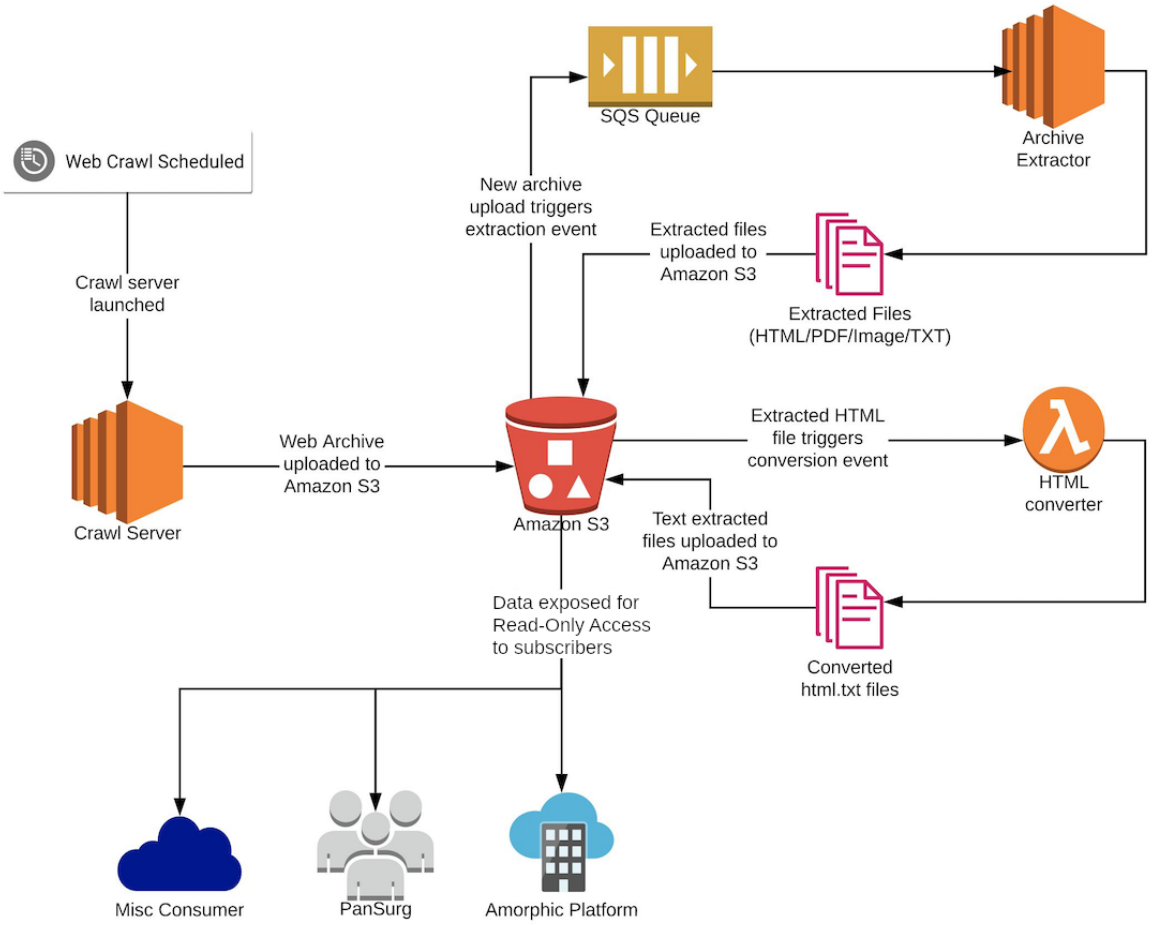
subfolder that contains relevant content, the web crawler will use a series of URL pattern matches and regular expressions to allow or disallow URL strings via exact matching or wildcard pattern matching, thereby containing the crawl to specific areas of the website. Additionally, scoping decisions can expand the crawl scope to include any outlying content that would be excluded by the refinement rules. Some websites have nonstandard locations for storing assets such as images, stylesheets, and scripts, which are needed to create a high-fidelity representation of the source material. Some websites also contain relevant documents that are located in prescribed location structures outside of the primary target folder.

Successful, high-fidelity targeted web crawling has been well documented. However, considerable challenges remain in the development of a qualitative and quantitative method for real-time relevant URL detection [9]. This is because the web ecosystem is a constantly evolving landscape with continuous advancements in available technology; construction techniques; and the consideration of desktop, mobile, and accessible display devices. Furthermore, the sheer number of content management systems that adopt their own proprietary content structures, the advent of single page applications, the prevalence of JavaScript, and people's dependence on asynchronous loading and POST method requests (which returns the same URL for multiple requests) render traditional URL similarity detection a particular challenge. Programmatic links with no human-readable semantic structures and features, such as "Latest News" sections within a web page, can often skew the results of link-based page relevance analyses.

These challenges are exacerbated in the REDASA system, which is required to target data in both academic and nonacademic sources without a guaranteed schema, dialect, or topic. Previously developed methodologies for addressing this issue have been used to apply anchor text similarity scores, content similarity scores, and URL prediction scores (which are based on a set of starting keywords) to seed data [10,11]. These scoring models promote and remove keywords based on the detection of commonalities in discovered crawl data. However, this approach presents several challenges because it relies on good starting URLs that present a reliable and consistent pattern of data. To counter this, REDASA performs downstream content filtering after the crawl is complete in order to eliminate extraneous data, which eliminates the risk of losing vital information that comes with the analysis of a potentially biased set of keywords. Due to REDASA's ability to perform a retrospective analysis of retained data, it will serve as a future platform that can be further enhanced by discovery automation.

In practice, REDASA performs an initial crawl that is launched by using crawl scope definitions that are governed by the aforementioned decision rules. The process for completing a crawl is outlined in Figure 1, which provides an interactive replica of the website content that is accessible to curators. MirrorWeb's Crawl Quality team reviewed the quality of the archive by using a replay engine to create a navigable replica of the target website archive. In addition to clicking the links and manually reviewing the content, an automated link checker was used to recursively spider the archive, identify page and document links in the HTML content, and attempt to open the target URLs in the archive instance in an effort to detect any missing referenced content. If any changes to the scoping rules are needed to make the web crawl more permissive or more restrictive, the scoping rules will be amended as required by using the same, aforementioned scoping principles. This iterative cycle repeats until the crawl is of sufficient quality to be used for human curation and accurate natural language processing (NLP) searches.

Figure 1. The REDASA back-end web crawling and data processing pipeline. REDASA: Realtime Data Synthesis and Analysis; SQS: Simple Queue Service; TXT: text.



Data Lake

Cloudwick's Amorphic platform provides a core REDASA data lake repository and data workflow orchestration service. MirrorWeb data initially lands in the storage layer of Amorphic, which consists of a landing zone. After validation checks are performed, data are moved and stored in a data lake zone and made available for document curation and search index workflows, as described in the following section.

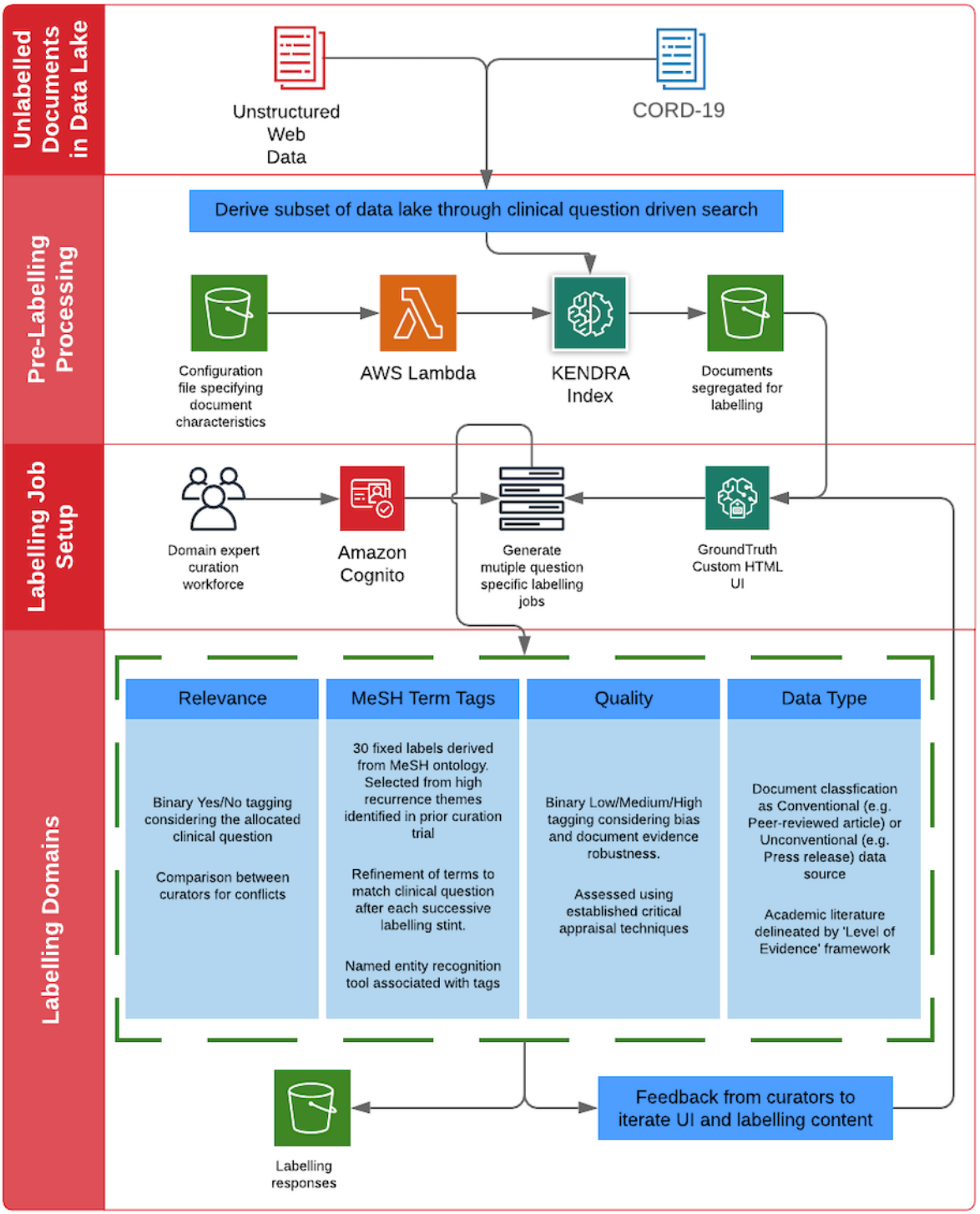
Search Index for Question-Specific Curation Documents

REDASA uses an AWS enterprise search service—Amazon Kendra—to provide search functionality across the entire data lake. Amazon Kendra is an NLP machine learning service that uses deep learning models to understand natural language queries and document content and structures. Amazon Kendra provides support for the following three broad types of questions: (1) factoid questions (who, what, when, and where), which are questions that require fact-based answers that may be returned in the form of a single word or phrase (the precise answer however must be explicitly stated in the ingested text content); (2) descriptive questions, which involves answers that could be a sentence, passage, or an entire document; and (3) keyword searches, wherein the intent and scope of the question may not be clear. In REDASA's question-specific curation model, Amazon Kendra exclusively received factoid questions and used a series of deep learning models to return relevant documents (Figure 2).

The key component of Amazon Kendra is an index. Conceptually, an index is an abstraction that encompasses a set of documents and the underlying hardware and software infrastructure that makes it possible to query documents that use natural language. Aside from its actual content, each document may include some associated metadata (eg, the source of the document, the document's logical unit group, etc). Users can specify custom metadata fields to suit their needs. These metadata tags are accessible through the Amazon Kendra query application programming interface.

A Kendra index may consist of one or more data sources, and a data source is a logical unit of documents. For REDASA, data source file types were limited to plain text, HTML, and PDF. Compared to other file types, these better integrate with our curation platform and allow for consistent labeling outputs when performing named entity recognition (NER).

Figure 2. Integrated workflow of the search index and data curation pipeline for a variety of high-impact areas with and without consensus among the scientific community in different countries and health authority bodies. AWS: Amazon Web Service; CORD-19: COVID-19 Open Research Dataset; MeSH: Medical Subject Headings; UI: user interface.



Document Curation

Document curation was implemented by using the custom workflows in Amazon SageMaker Ground Truth, which is a data labeling service that is used to build training data sets for machine learning workflows. REDASA uses a question-based curation approach. PanSurg investigators posed a series of COVID-19–related key questions to the search index (Textbox 1). These questions were chosen to obtain answers, and we were able to validate the quality of the data lake and the adequacy of REDASA's data mining depth with our curation relevance metric.

Textbox 1. COVID-19–related natural language processing queries that were posed to the REDASA (Realtime Data Synthesis and Analysis) search index to develop a question-specific curation methodology.

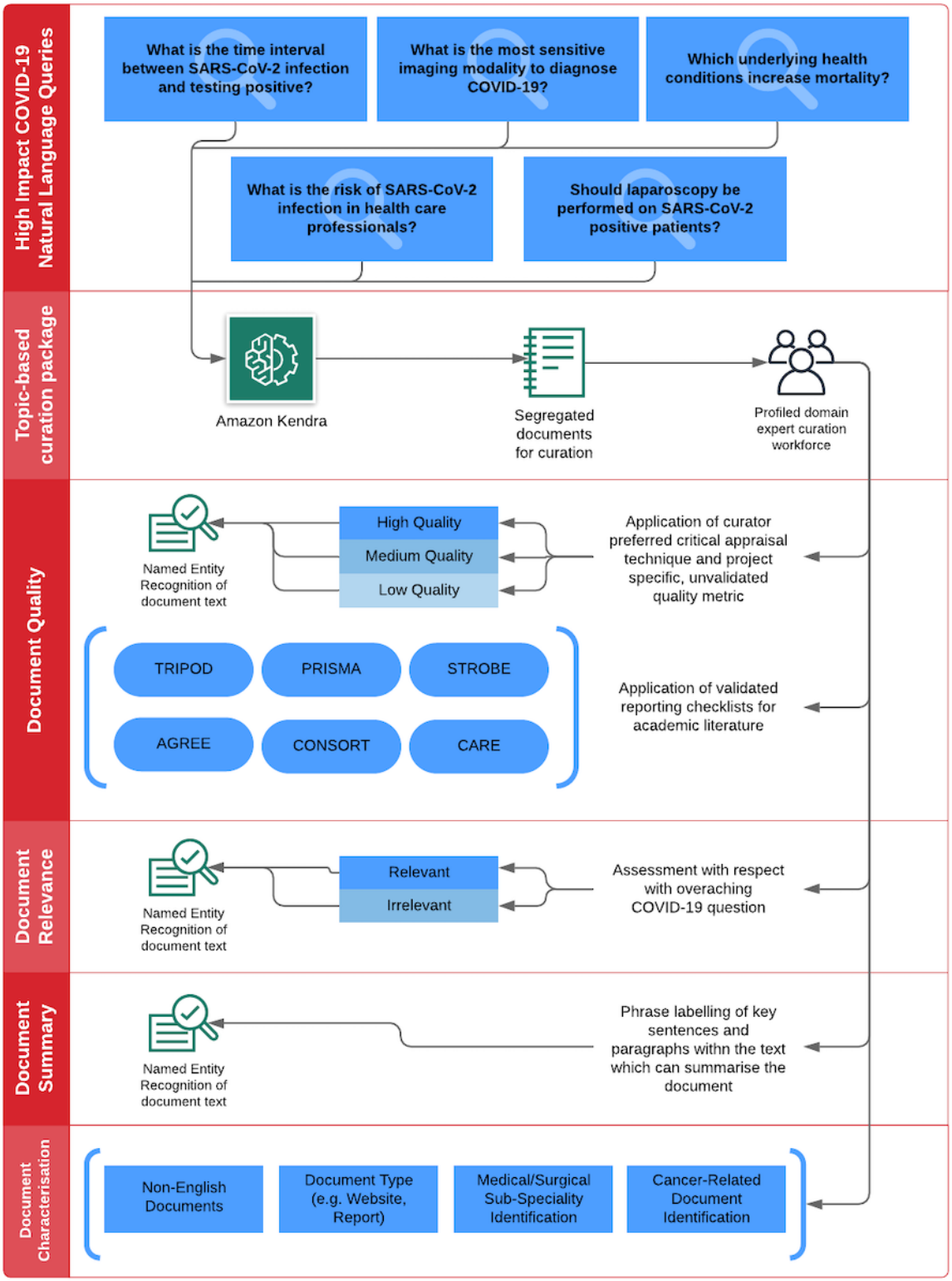
Queries in natural language

1. What is the time interval between SARS-CoV-2 infection and testing positive?
2. What is the most sensitive imaging modality to diagnose COVID-19?
3. Which underlying health conditions increase mortality?
4. What is the risk of SARS-CoV-2 infection in health care professionals?
5. Should laparoscopy be performed on SARS-CoV-2 positive patients?
6. What is the estimated global economic impact of COVID-19?
7. How effective are masks at minimizing the transmission of COVID-19 in public?
8. What is the evidence for COVID-19 mutations and how many subtypes are responsible for the pandemic?
9. Does a positive SARS-CoV-2 antibody test mean an individual is immune to further COVID-19 infections?
10. Is COVID-19 airborne transmitted?
11. Can asymptomatically infected individuals transmit COVID-19?
12. What is the evidence for 1-meter and 2-meter separations for social distancing?
13. What has the evidence-base been for lockdown compared to no lockdown during this COVID-19 pandemic?
14. Is universal preoperative testing for SARS-CoV-2 beneficial compared to selective testing?
15. Can individuals be reinfected with SARS-CoV-2?

The REDASA search index provides a list of selected documents, which are randomly provisioned to curators for labeling via Amazon SageMaker Ground Truth. This allows them to assess documents' relevance and quality in relation to the original query and further categorize the data based on the labels described in Figure 3. Reflecting the living nature of the REDASA platform, queries were adapted in accordance with the knowledge priorities of different phases of the pandemic. For example, in stint 1 of curation (February 6 to September 6, 2020), which was scheduled during the peak of the UK COVID-19 outbreak and when uncertainties regarding best practices for screening and management planning were rife, questions 1-5 were posed to REDASA. In contrast, stint 2 of curation was performed during the nationwide lockdown relaxation period and the public health transition for minimizing the risk of a second wave of COVID-19. Consequently, questions 6-15 focused upon themes such as reinfection, transmission mitigation, and the global impact of the pandemic.

The relevance of articles in relation to the query that was posed to the search index was a subjective binary measure (ie, irrelevant or relevant). This assessment was paired with NER labels, which enabled curators to highlight phrases and paragraphs that indicated the relevance of articles.

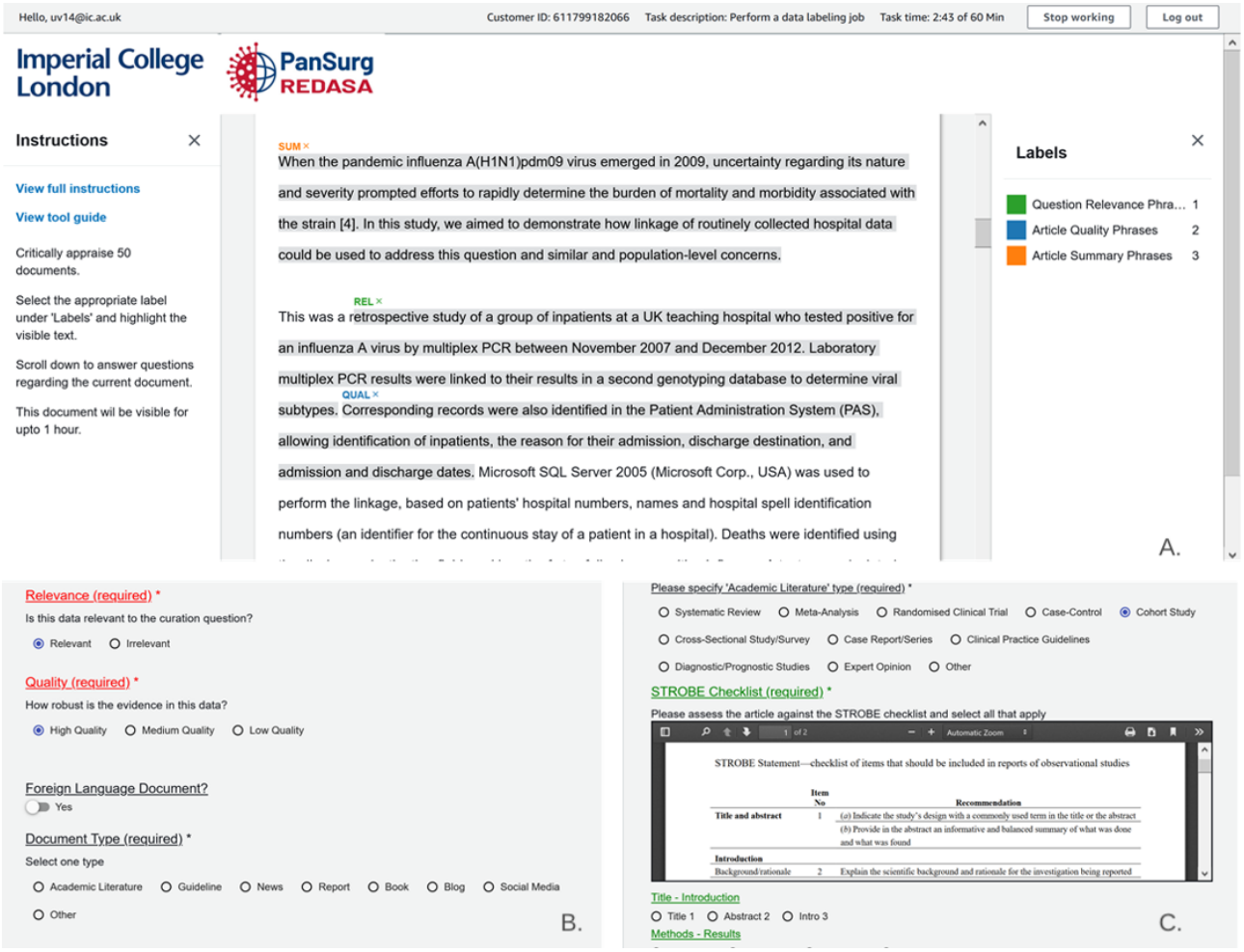
Figure 3. Curation labels for generating document metadata. AGREE: Appraisal of Guidelines for Research and Evaluation; CARE: Case Reports; CONSORT: Consolidated Standards of Reporting Trials; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis.



The quality of the academic literature was assessed via a 3-stage process. First, we ascertained the study type, and this allowed us to assign an evidence rating level (the levels proposed by the Oxford Centre for Evidence-Based Medicine [12]). Second, we invited curators to provide an independent, subjective rating of an article's quality by using their own critical appraisal methodology and assign 1 of the 3 following binary ratings: low, medium, or high. Third, akin to the relevance metric, NER annotation was made available to curators and correlated with their low, medium, or high ratings. Depending on the type of academic literature that curators were assessing, curators were automatically given (through their user interface) the relevant EQUATOR (Enhancing the Quality and Transparency of Health Research) checklist for quantitative quality assessment [13].

For example, if the document that curators were assessing was a systematic review, they were automatically able to assess the article against the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist (Figure 4).

Figure 4. (A) A document with curation user interface labels (the NER of quality, relevance, and summary phrases). (B) Binary labels for classifying documents and correlating them to NER responses. (C) Embedded reporting checklists for document assessment, which were provided based on the selected academic study type. NER: named entity recognition; REDASA: Realtime Data Synthesis and Analysis; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology.



Collectively, the relevance and quality metrics' utility was threefold. First, they enabled us to capture data on curators' decision-making and critical appraisal processes. Second, they minimized the number of undesirable irrelevant documents, which allowed us to implement a human-in-the-loop optimization methodology for the search index. Third, they allowed us to perform multiple curator passes on a single document, assess for labeling response conflicts, and ascertain the article factors responsible for any disparities. To obtain further data, we allowed our curators to assess the risk of bias in the articles by using the bias metrics designed by Cochrane [14]. The results from this novel curation process were intended to (1) act as ground truth for data science models that aimed to facilitate the future semiautomation or full automation of article screening, and (2) be used for a structured assessment of evidence quality.

This question-specific approach was selected over the more traditional approach of randomly sectioning data to help us preserve the relevance metric for specific questions. This factor would have otherwise been more challenging to implement and capture. Further, this metric is key to future work streams that determine the relevance of specific articles for inclusion in an automated systematic review.

Curation Methodology

Structured and Unstructured Data Lake

Our proof-of-concept analysis for data mining during predefined time periods was feasible. In this iteration of REDASA, a 1-week time period was chosen to enable the capturing of the highest possible number of new data points at the lowest mining frequency, thus minimizing the computing costs of both COVID-19–structured and unstructured data sources. In this paper, we only present a text-based analysis of the data set. In the future, we intend to assess structured, quantitative data from target data sources.

Unstructured Web Crawl Data

Textual information was extracted from a precurated set of internet sites. A range of frequently accessed but disparate data types that are typically used by frontline clinicians and policy makers were extracted. These included high-quality journal websites and portals containing COVID-19–related literature;

medical and surgical society guidance web pages; and guideline repositories from local, governmental, and international public health bodies. By being able to dynamically capture data and automatically obtain updates from sources, this type of data mining demonstrates the power of REDASA in terms of amalgamating qualitative and quantitative insights for generating future reports. Each website was independently assessed and evaluated for inclusion into the REDASA data lake by clinical (n=4) and data science (n=2) reviewers. Disagreements were resolved through consensus. These sources were selected in accordance with criteria for including usable content and determining the reliability and breadth of target topics and categories pertaining to COVID-19. To systematize the data lake prior to data ingestion, sources were categorized into the following broad groups, which were independently defined a priori by the three members of the research team based on the source of the original data: all (miscellaneous), critical care, medical care, surgical care, drug development and pharmacological therapy, mental health, risk, translational research, biological sciences, engineering, and policy. The content of the data from each of the sources was screened by these three independent members of the research team. If disagreements regarding categorization occurred, a meeting was conducted. Unanimous agreement was sought prior to final categorization.

Structured Data From the COVID-19 Open Research Dataset

The White House and a coalition of leading research groups developed the COVID-19 Open Research Dataset (CORD-19) [15]. The CORD-19 is a data set that contains over 157,000 scholarly articles, including over 75,000 full-text articles regarding COVID-19, SARS-CoV-2, and related coronaviruses. This freely available data set was provided to the global research community of Kaggle and was used as a test data source when initially developing the REDASA infrastructure.

Collectively, these assimilated data sources make REDASA one of the world's largest and most contemporaneous COVID-19–related evidence sources, consisting of 104,000 documents.

Data Availability

The curation labels can be found on GitHub [16].

Results

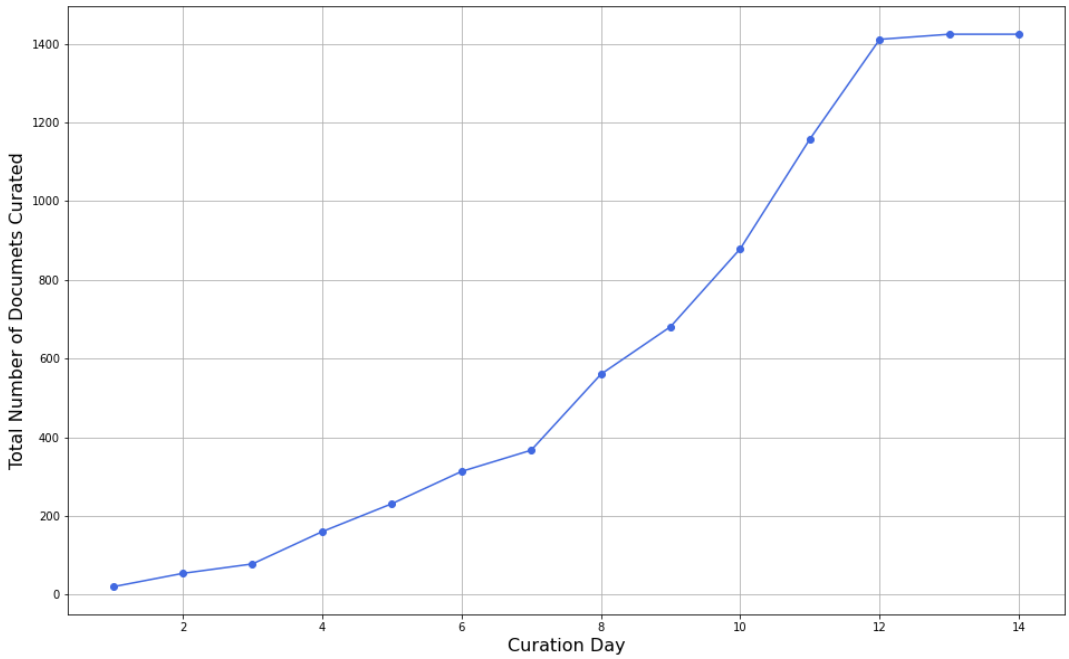
Curation Results

The first-pass document curation responses from 42 curators underwent the preliminary analysis of the domains of relevance to free-text queries and information quality. These data were collected over 2 time-limited curation stints. Both stints were 1 week in duration. We obtained a total of 1424 documents pertaining to 15 different COVID-19–related queries (around 99 text documents derived per query from the search index for the structured and unstructured data lake), and an average of 42 documents per query were assessed by 1 curator. Our aim was for each document to be assessed by at least 2 different reviewers so that we could assess intercurator variability and identify the reasons for discrepancies in evidence quality verdicts.

Each curator was profiled to ascertain their academic or clinical backgrounds. This was initially performed for the vetting and quality control of curation responses. These data will also be used to further discriminate between labeling responses that are based on critical appraisal expertise and to assign weights to curation responses. To date, the REDASA project's international curator community includes people from 9 different countries, and the project is supported by medical and surgical health care professionals who range from senior consultant vascular surgeons in Italy to general practice physicians who are involved in community and public health decision making in the Philippines. Data curators were recruited by invitation through the PanSurg email subscriber list and by open invitation via the @pansurg Twitter account. Curators were included if they had a medical degree or a higher degree in science-related fields. Data curators were asked to state their interest and were verified by direct contact. The number of data curation responses for REDASA, which exponentially rose between our two stints as more curators were onboarded (stint 1: n=12; stint 2: n=42; Figure 5), was indicative of an efficient, novel methodology for the digital, community-based peer review of literature by domain experts. This was exemplified by the ability of some of our experienced curators, who were able review over 100 scientific documents of varying length in as little as 3 days. Furthermore, with curators' 100% follow-through rate between stint 1 and stint 2, our curation model suggests that, when combined with our simplified critical appraisal interface, the peer review of literature at scale is viable and sustainable.

In total, 70.9% (1009/1424) of the pool of curated articles was composed of peer-reviewed, traditional, academic literature; the remainder consisted of web crawl–derived data, including governmental policies and reports from professional bodies. Based on the subset of the 900 academic literature documents that were curated, the most common study type encountered was systematic reviews (98/1009, 9.7%). The least common study type in the data lake was randomized controlled trials (RCTs; 3/1009, 0.3%). Nonsystematic reviews (eg, rapid reviews and comprehensive reviews) were not given the systematic review label to avoid inappropriate assessments against the PRISMA checklist. Such outlier academic literature types were aggregated into the miscellaneous category, which included 427 documents that were assessed solely against the curator-reported binary quality ratings.

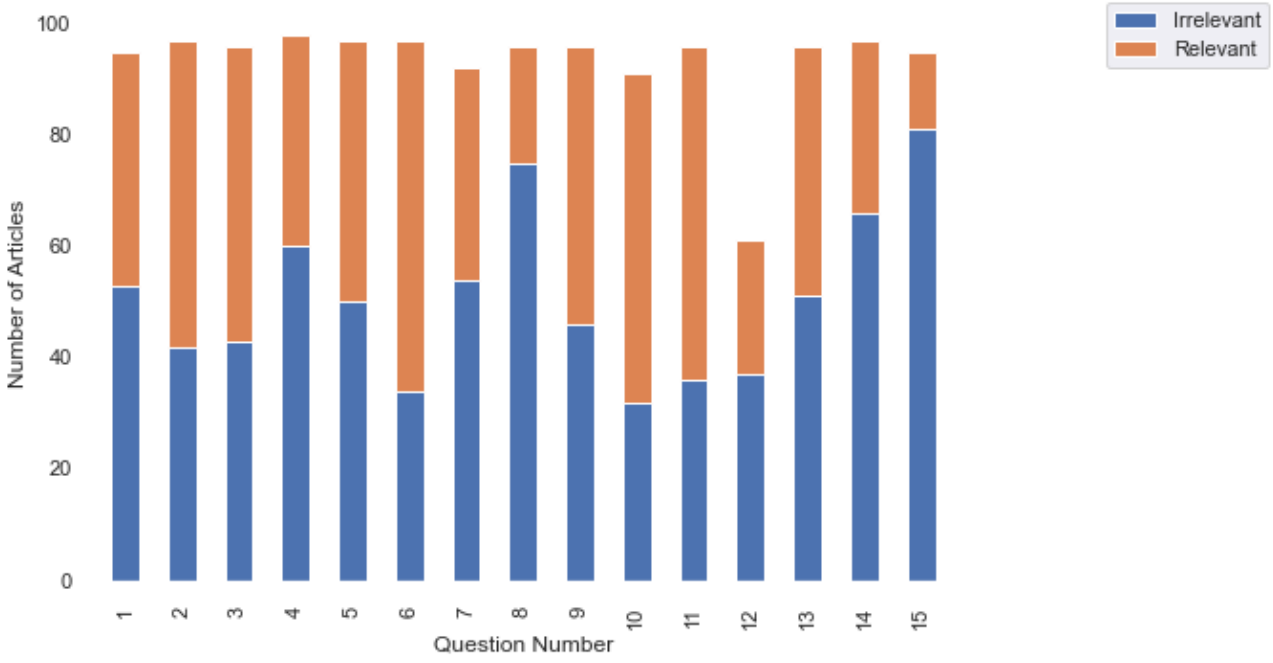
Figure 5. Rate of COVID-19–related scientific literature curation over 2 weeks. This was associated with the growth of the number of curators, which plateaued on day 13. This was when all of the documents available for curation were assessed before the end of stint 2.



Articles' Relevance to Queries

The relevance metric that curators used provided insight into the performance of the search index in terms of providing cogent and useful document results associated with the 15 COVID-19–related queries (Textbox 1). Overall, 50.49% (719/1424) of articles were considered relevant to their respective queries. When observing the question bank, this variance in article relevance (which was based on the search index) was reflected by the lack of consistency in the ratio of the number of relevant articles to the number of irrelevant articles (Figure 6). These data can be used to provide feedback for the search index with regard to the optimization of provided results.

Figure 6. Curators' responses determined the relevance of documents to search index queries. Responses were matched to the query number.



Critical Appraisal of Article Quality

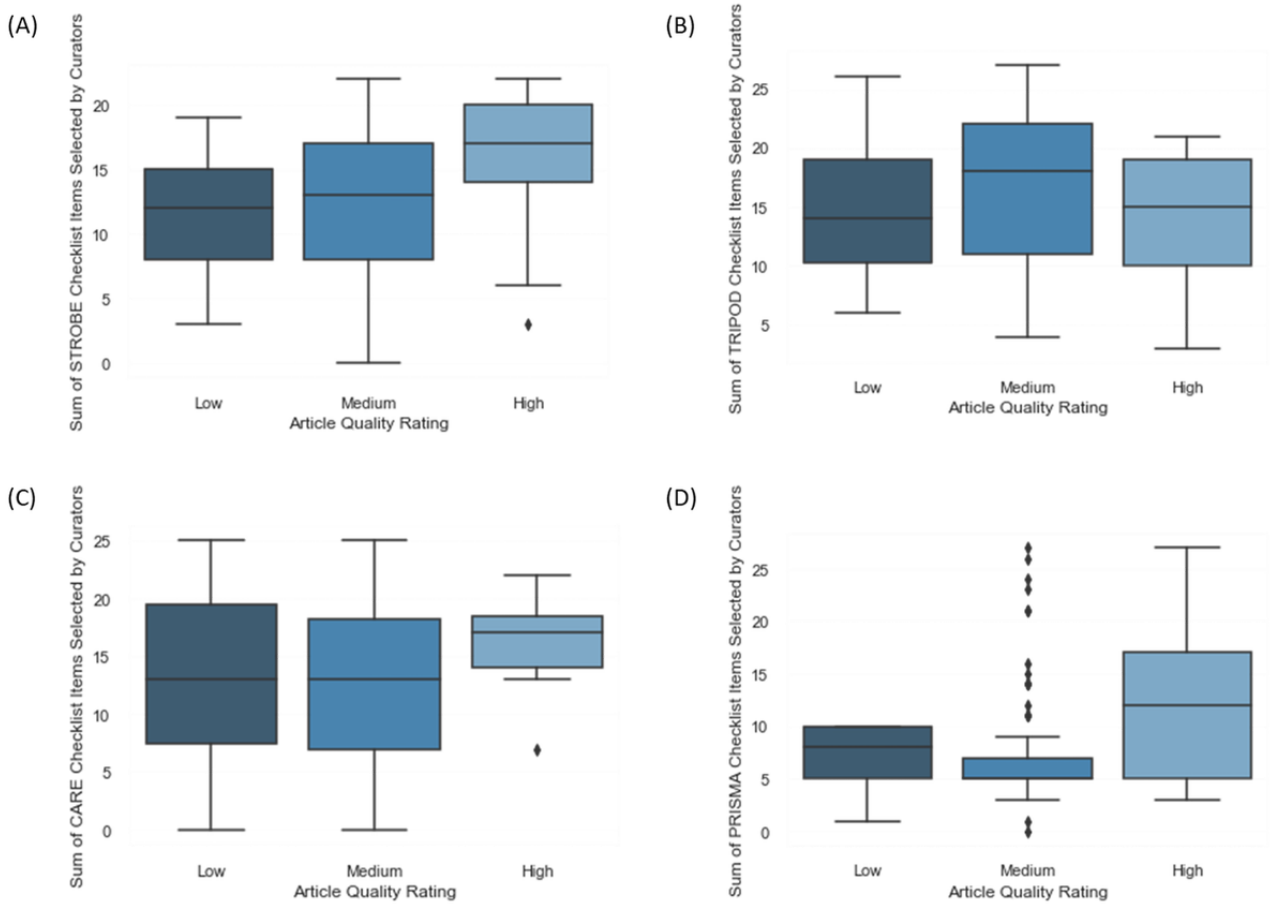
The uneven distribution of academic study types that have been curated thus far precludes the interpretation of results for quantitative reporting checklist responses based on the qualitative rating system (low, medium, and high) for RCTs (CONSORT [Consolidated Standards of Reporting Trials] checklist) and clinical guidelines (AGREE [Appraisal of Guidelines for Research and Evaluation] checklist). These studies were poorly represented in this run of analysis.

Quality was quantified by ascertaining the sum of the number of EQUATOR Network–derived checklist items that were fulfilled by each document. Hence, documents with methods and results that aligned more closely to their respective reporting checklist were scored higher and deemed to be of greater quality. This outcome was compared to the curators' subjective ratings for diagnostic and prognostic studies (TRIPOD [Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis] checklist), case reports and series (CARE [Case Reports] checklist), case-control studies (STROBE [Strengthening the Reporting of Observational Studies in Epidemiology] checklist) and meta-analyses and systematic reviews (PRISMA checklist) (Figure 7). Notably, the subjective quality rating was assigned prior to assessment by using the checklist under our curation protocol to mitigate observer bias.

Based on our independently assessed and subjective quality metric, our preliminary results suggested that more than 50% (726/1424, 50.98%) of the documents derived from the REDASA data pipeline were of medium quality, and 13.55% (193/1424) were deemed high quality. Thus, during data aggregation, 64.53% (919/1424) of the documents derived from our data lake for curation were of a sufficient quality to inform their professional decision-making processes or could be used as reliable sources of information.

With regard to the TRIPOD-relevant (score: mean 15.6, SD 6.9) and STROBE-relevant (score: mean 13.2, SD 5.8) study types, there was some correlation between curators' subjective assessments of the articles that had low-to-high ratings and underwent the validated checklist–based quality assessment. However, with regard to the CARE-relevant (score: mean 13.8, SD 6.8) and PRISMA-relevant (score: mean 8.4, SD 6.1) study types, there was substantial variance in the number of checklist items that were selected for each quality rating, thus indicating an apparent dissociation between these two metrics. Further data collection and the comparison of intercurator responses for the removal of outliers will provide clarity on the role of subjective quality ratings versus the role of validated reporting checklists as a surrogate marker of evidence quality.

Figure 7. Relationship between the low, medium and, high curator-determined quality ratings of (A) case-control studies, (B) diagnostic and prognostic studies, (C) case reports and series, and (D) meta-analyses and systematic reviews and their respective reporting checklist scores. CARE: Case Reports; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis.



Discussion

Globally, there are several efforts underway for systematically accruing COVID-19–related data and specifically querying these data and output-relevant literature [17]. These efforts, such as Google's COVID-19 Research Explorer [18], COVIDask [19], and COVIDScholar [20], are fundamentally based on NLP searches. Additionally, Google's solution incorporates the use of follow-up queries associated with a primary question to obtain more focused results. These efforts universally incorporate the CORD-19 and intuitively present output data. Nevertheless, these approaches do not account for the quality of the data source, which is left to the interpretation of the user. Other efforts, such as SciSight [21] and COVID-19 Primer [22], structure data from the CORD-19 into themes (author and engagement), thereby allowing users to make links and review specific topics, albeit without a natural language interface for answering specific questions.

The crucial difference between REDASA and the aforementioned platforms is threefold. First, REDASA adopts a human-in-the-loop methodology by embedding an efficient, user-friendly curation platform into an NLP search engine. REDASA can iteratively refine its search outputs at scale, particularly in the domains of the relevance and quality of data sources. This can ultimately contribute to a fact-checking function for conducting a reliable assessment of the utility of an article [23]. Second, it provides a curated data set in the JavaScript Object Notation format for experienced academic reviewers' critical appraisal choices and decision-making methodologies. These data on the peer-review process provide a unique framework for modelling, quantifying, and ultimately automating the evidence quality assurance process and are unavailable elsewhere. Finally, due to the wide scope and depth of REDASA's web crawling methodology, REDASA has already captured one of the world's largest COVID-19–related data corpora for searches and curation. Our aim is to make these crucial data freely available and ensure that they are continuously updated to allow for rapid review and dissemination during and beyond the evolving pandemic.

For the long-term goal of conducting a semisupervised, live systematic review of data, several limitations and challenges need to be overcome. Our curation methodology resulted in a high turnover rate for the assessment of data. However, there was still variability in curator output, which was secondary to the variability in curators' subjective critical appraisals. In this project, we relied on the prescreening of curators, which was conducted via academic portfolio screening and assessments for relevant literature review experience. This crucial quality control approach needs to be further developed to fully validate and enhance the accuracy of our curation methodology. A limitation of our preliminary data analysis was the qualitative, summative comparison of the EQUATOR checklist ratings to our quality ratings. This was due to the subcomponents of the used EQUATOR checklists, which did not use equal metrics for article quality, and the nonexhaustive quality criteria captured by these tools. Hence, future studies are needed to validate our quality ratings and identify a reliable metric for quality that is applicable across the academic and nonacademic literature captured by REDASA. In addition to ensuring the consistency of quality ratings, sustained curation work is required to ensure that the corpus includes greater numbers of studies across all designs and methodologies—specifically, RCTs (if available)—to ensure that the corpus is truly representative of data under examination.

Our framework has demonstrated proof-of-concept that by combining the discovery and ingestion pipeline, data lake repository, human curation platform, and NLP semantic search index of REDASA, it can provide curated responses to questions that are posed in natural language in the short term. In the long term however, based on the data insights that progressively validated the critical appraisal of our curation methodology, the ambition of REDASA is to conduct live systematic reviews by using semisupervised machine learning techniques to rapidly return high-quality, relevant evidence in response to queries for any discipline experiencing an "infodemic," such as cancer or cardiovascular disease.

Acknowledgments

We would like to thank the team at AWS Healthcare for their support and help throughout the development of this project and Simon Russell and Steve Neufville for their technical and project management support, respectively. We would also like to thank the PanSurg Collaborative for their support during the project. We thank David Hilman for his technical support with the final stages of this paper. We also thank Elizabeth Burgnon for carefully proofreading the final paper. This work was supported by Defence and Security Accelerator (grant ACC2015551), the Digital Surgery Intelligent Operating Room Grant, the National Institute for Health Research Long-limb Gastric Bypass RCT Study, the Jon Moulton Charitable Trust Diabetes Bariatric Surgery Grant, the National Institute for Health Research (grant II-OL-1116-10027), the National Institutes of Health (grant R01-CA204403-01A1), Horizon H2020 (ITN GROWTH), and the Imperial Biomedical Research Centre.

The members of the PanSurg REDASA Curators are Ademola Adeyeye (Afe Babalola University), Ahmed Ezzat (Imperial NHS Trust), Alberto Porcu (Università degli Studi di Sassari, Italia), Alexander Walmsley (University of Liverpool), Ali Farsi (King Abdulaziz University, Jeddah, Saudi Arabia), Alison Faye O Chan (Institute of Biology, College of Science, University of the Philippines Diliman), Aminah Abdul Razzack (Dr NTR University of Health Sciences), Andee Dzulkarnaen Zakaria (School of Medical Sciences, Universiti Sains Malaysia), Andrew Yiu (Guy's and St Thomas' NHS Foundation Trust), Antonios Soliman (Assiut University Hospitals), Ariana Axiaq (Queen's University Belfast), Avinash Aujayeb (Northumbria HealthCare NHS Foundation Trust), Catherine Dominic (Barts and the London School of Medicine, Queen Mary University of London), Eduarda Sá-Marta (Faculty of Medicine, University of Coimbra), Eunice F Nolasco (Independent Researcher), Jessamine Edith S Ferrer

(Blessed Family Doctors General Hospital), Jonathan Anthony Kat (University of Manchester), Josephine Holt (Chelsea and Westminster NHS Trust), Kamal Awad (Faculty of Medicine, Zagazig University, Egypt), Kirk Chalmers (University of Liverpool), Mina Ragheb (University of Liverpool School of Medicine), Muhammad Khawar Sana (King Edward Medical University), Niraj Sandeep Kumar (UCL Medical School), Roland Amoah (Newcastle upon Tyne Hospitals Trust), Semra Demirli Atici (University of Health Sciences Tepecik Training and Research Hospital), Shane Charles (San Fernando General Hospital), Sunnia Ahmed (Independent Researcher), Teresa Perra (Università degli Studi di Sassari, Italia), Tricia Hui Chyi TAY (Manchester Medical School, University of Manchester), Ubaid Ullah (Medical Emergency Resilience Foundation Pakistan), Zara Ahmed (King's College London), and Zun Zheng Ong (University of Nottingham).

Authors' Contributions

JK, GM, and MH were responsible for project conception and overall management. The development of the data pipeline and data lake structure was steered by KS, EF, HS and DB. The implementation of the MirrorWeb web crawling technology was led by KS, who received guidance from UV, SR, PB, and OS. The initial development of the curation platform was performed by EF, and subsequent optimization and tailoring for REDASA use was conducted by CL, UV, SR, and OS. The coordination and development of the data curation methodology was performed by UV, SR, PB, and OS. Data analysis and manuscript composition was performed by UV and SR, who received input from JK, GM, OS, SP, JC, KS, EF, and MH. All authors reviewed and contributed to the manuscript. We would like to thank the PanSurg REDASA Curators for their support on the project.

Conflicts of Interest

GM received equity from Medical iSight (Augmented Reality). SP provides consultations for Medtronic, T.M.L.E. Ltd., and Roche. SP is also the cofounder and director of Mangetoo, 1 World Medical, and the London General Surgery Clinic. SP is also a partner of One Welbeck Hospital. JK provides consultations for Verb robotics, Safeheal, YSOPIA bioscience, and Universal Diagnostics (UDX). JK also received equity from Mangetoo (teledietetics), 1 Welbeck Day Surgery (Hospital), 1 World medical (Personal Protective Equipment), and Medical iSight (Augmented Reality). The other authors have no conflicts of interest to declare for this paper.

References

1. LitCovid. National Center for Biotechnology Information. URL: https://www.ncbi.nlm.nih.gov/research/coronavirus/ [accessed 2021-04-12]
2. COVID-19 research update: How many pandemic papers have been published? Nature Index. URL: https://www.natureindex.com/news-blog/how-coronavirus-is-changing-research-practices-and-publishing [accessed 2020-10-01]
3. COVID-19 rapid guideline: critical care in adults. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/ng159 [accessed 2020-10-01]
4. Interim process and methods for developing rapid guidelines on COVID-19. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/process/pmg35/chapter/scoping [accessed 2020-10-01]
5. Countering Missinformation about COVID-19. World Health Organization. URL: https://www.who.int/news-room/feature-stories/detail/countering-misinformation-about-covid-19 [accessed 2020-10-01]
6. Maguire BJ, Guérin PJ. A living systematic review protocol for COVID-19 clinical trial registrations. Wellcome Open Res 2020 Apr 02;5:60 [FREE Full text] [doi: 10.12688/wellcomeopenres.15821.1] [Medline: 32292826]
7. Elliott JH, Turner T, Clavisi O, Thomas J, Higgins JPT, Mavergames C, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. PLoS Med 2014 Feb;11(2):e1001603 [FREE Full text] [doi: 10.1371/journal.pmed.1001603] [Medline: 24558353]
8. Lenaerts G, Bekkering GE, Goossens M, De Coninck L, Delvaux N, Cordyn S, et al. Tools to Assess the Trustworthiness of Evidence-Based Point-of-Care Information for Health Care Professionals: Systematic Review. J Med Internet Res 2020 Jan 17;22(1):e15415 [FREE Full text] [doi: 10.2196/15415] [Medline: 31951213]
9. Olston C, Najork M. Web Crawling. Boston, United States: Now Publishers Inc; 2010.
10. Rungsawang A, Angkawattanawit N. Learnable topic-specific web crawler. Journal of Network and Computer Applications 2005 Apr;28(2):97-114. [doi: 10.1016/j.jnca.2004.01.001]
11. Zowalla R, Wetter T, Pfeifer D. Crawling the German Health Web: Exploratory Study and Graph Analysis. J Med Internet Res 2020 Jul 24;22(7):e17853 [FREE Full text] [doi: 10.2196/17853] [Medline: 32706701]
12. Oxford Centre for Evidence-Based Medicine: Levels of Evidence. Centre for Evidence-Based Medicine. URL: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009 [accessed 2021-04-22]
13. Enhancing the QUAlity and Transparency Of health Research. The Equator Network. URL: https://www.equator-network.org/ [accessed 2021-04-22]
14. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions, Second Edition. Glasgow, Scotland: The Cochrane Collaboration and John Wiley & Sons; Nov 20, 2019.

15. Lu Wang L, Lo K, Chandrasekhar Y, Reas R, Yang J, Eide D, et al. CORD-19: The Covid-19 Open Research Dataset. ArXiv Preprint posted online on April 22, 2020. [Medline: 32510522]
16. PanSurg/redasa-sample-data. GitHub. URL: https://github.com/PanSurg/redasa-sample-data [accessed 2021-04-26]
17. Hutson M. Artificial-intelligence tools aim to tame the coronavirus literature. Nature 2020 Jun 9 [FREE Full text] [doi: 10.1038/d41586-020-01733-7]
18. COVID-19 Research Explorer. Google. URL: https://covid19-research-explorer.appspot.com/ [accessed 2021-04-22]
19. COVIDAsk. DMIS Lab, Korea University. URL: https://covidask.korea.ac.kr/ [accessed 2020-10-01]
20. COVIDScholar. COVIDScholar. URL: https://covidscholar.org/ [accessed 2021-04-22]
21. SciSight - About. SciSight. URL: https://scisight.apps.allenai.org/ [accessed 2021-04-22]
22. COVID-19 Primer. Primer.ai. URL: https://covid19primer.com/dashboard [accessed 2021-04-22]
23. Wadden D, Lin S, Lo K, Wang LL, van Zuylen M, Cohan A, et al. Fact or fiction: Verifying scientific claims. arXiv Preprint posted online on April 30, 2020. [doi: 10.18653/v1/2020.emnlp-main.609]

Abbreviations

AGREE: Appraisal of Guidelines for Research and Evaluation
AWS: Amazon Web Service
CARE: Case Reports
CONSORT: Consolidated Standards of Reporting Trials
CORD-19: COVID-19 Open Research Dataset
EQUATOR: Enhancing the Quality and Transparency of Health Research
NER: named entity recognition
NLP: natural language processing
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
REDASA: Realtime Data Synthesis and Analysis
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis

Edited by C Basch; submitted 12.11.20; peer-reviewed by S Liu, B Sousa-Pinto; comments to author 11.12.20; revised version received 30.12.20; accepted 03.04.21; published 06.05.21

Please cite as:
Vaghela U, Rabinowicz S, Bratsos P, Martin G, Fritzilas E, Markar S, Purkayastha S, Stringer K, Singh H, Llewellyn C, Dutta D, Clarke JM, Howard M, PanSurg REDASA Curators, Serban O, Kinross J
Using a Secure, Continually Updating, Web Source Processing Pipeline to Support the Real-Time Data Synthesis and Analysis of Scientific Literature: Development and Validation Study
J Med Internet Res 2021;23(5):e25714
URL: https://www.jmir.org/2021/5/e25714
doi: 10.2196/25714
PMID: 33835932

Review

Machine Learning and Natural Language Processing in Mental Health: Systematic Review

Aziliz Le Glaz[1], MD; Yannis Haralambous[2], PhD; Deok-Hee Kim-Dufor[1], PhD; Philippe Lenca[2], PhD; Romain Billot[2], PhD; Taylor C Ryan[3], MS; Jonathan Marsh[4], MA; Jordan DeVylder[4], MD, PhD; Michel Walter[1,5], MD, PhD; Sofian Berrouiguet[1,2,5,6], MD; Christophe Lemey[1,2,5], MD

[1]URCI Mental Health Department, Brest Medical University Hospital, Brest, France

[2]IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238, Brest, France

[3]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[4]Fordham University Graduate School of Social Service, New York, NY, United States

[5]EA 7479 SPURBO, Université de Bretagne Occidentale, Brest, France

[6]LaTIM, INSERM, UMR 1101, Brest, France

Corresponding Author:
Christophe Lemey, MD
URCI Mental Health Department
Brest Medical University Hospital
Route de Ploudalmézeau
Brest, 29200
France
Phone: 33 619211032
Email: christophe.lemey@chu-brest.fr

Abstract

Background: Machine learning systems are part of the field of artificial intelligence that automatically learn models from data to make better decisions. Natural language processing (NLP), by using corpora and learning approaches, provides good performance in statistical tasks, such as text classification or sentiment mining.

Objective: The primary aim of this systematic review was to summarize and characterize, in methodological and technical terms, studies that used machine learning and NLP techniques for mental health. The secondary aim was to consider the potential use of these methods in mental health clinical practice

Methods: This systematic review follows the PRISMA (Preferred Reporting Items for Systematic Review and Meta-analysis) guidelines and is registered with PROSPERO (Prospective Register of Systematic Reviews; number CRD42019107376). The search was conducted using 4 medical databases (PubMed, Scopus, ScienceDirect, and PsycINFO) with the following keywords: machine learning, data mining, psychiatry, mental health, and mental disorder. The exclusion criteria were as follows: languages other than English, anonymization process, case studies, conference papers, and reviews. No limitations on publication dates were imposed.

Results: A total of 327 articles were identified, of which 269 (82.3%) were excluded and 58 (17.7%) were included in the review. The results were organized through a qualitative perspective. Although studies had heterogeneous topics and methods, some themes emerged. Population studies could be grouped into 3 categories: patients included in medical databases, patients who came to the emergency room, and social media users. The main objectives were to extract symptoms, classify severity of illness, compare therapy effectiveness, provide psychopathological clues, and challenge the current nosography. Medical records and social media were the 2 major data sources. With regard to the methods used, preprocessing used the standard methods of NLP and unique identifier extraction dedicated to medical texts. Efficient classifiers were preferred rather than transparent functioning classifiers. Python was the most frequently used platform.

Conclusions: Machine learning and NLP models have been highly topical issues in medicine in recent years and may be considered a new paradigm in medical research. However, these processes tend to confirm clinical hypotheses rather than developing entirely new information, and only one major category of the population (ie, social media users) is an imprecise cohort. Moreover, some language-specific features can improve the performance of NLP methods, and their extension to other languages should be more closely investigated. However, machine learning and NLP techniques provide useful information from

unexplored data (ie, patients' daily habits that are usually inaccessible to care providers). Before considering It as an additional tool of mental health care, ethical issues remain and should be discussed in a timely manner. Machine learning and NLP methods may offer multiple perspectives in mental health research but should also be considered as tools to support clinical practice.

(J Med Internet Res 2021;23(5):e15708) doi: 10.2196/15708

KEYWORDS

machine learning; natural language processing; artificial intelligence; data mining; mental health; psychiatry

Introduction

Machine Learning

Machine learning (ML) systems automatically learn models from data to make better decisions. As such, they are part of a major subfield of artificial intelligence (AI). There are 3 main approaches to learning from data: supervised, unsupervised, and reinforcement learning. In supervised learning, a target attribute is predicted, and ML algorithms infer a model from labeled input data (ie, a training data set that provides examples described by predictive attributes and values for the target attribute). The goal is to make target predictions on new data to obtain good generalization performance. In contrast, there is no target attribute in unsupervised learning, and thus no labeled data. Unsupervised learning consists of inferring a model to describe hidden patterns from unlabeled data. Under circumstances in which labeled data acquisition proves to be difficult, (eg, costly), semisupervised ML methods can use both labeled and unlabeled data for learning. The third main category of ML is reinforcement learning, in which the ML model uses feedback that acts as a reward or punishment to maximize its performance.

ML is limited to certain capacities. For one, it relies on collections of data that may be incomplete, noisy, or subject to systematic bias, all of which can lead to erroneous predictions. In addition, ML algorithms may introduce bias. Interesting questions to be addressed in ML are discussed in an article by Domingos [1]. However, when carefully conducted, ML can have great utility.

AI and ML have many applications, many of which are encountered in daily life. Supervised ML, for example, is widely used for spam filtering (ie, classifying incoming email as spam or not spam) [2]. It is also used to classify credit applicants based on their probabilities of default [3]. Unsupervised ML, such as algorithm clustering, is able to group customers with similar characteristics and their likelihood to purchase. This is widely used by banks for market segmentation [4]. Finally, automatic document clustering that organizes similar documents into classes (for purposes of improving information retrieval, for example) is gaining importance due to the increasing number of documents on the internet [5].

The application of ML in health is also of concern. Indeed, ML is widely used in critical disease models in cardiology, neurology, and diabetes research [6] to automatically identify heart disease risk factors [7], to classify primary progressive aphasia subtypes [8], and for the characterization and diagnosis of cognitive impairments [9], diabetes, and cardiovascular disorders [10-17].

ML is also challenging the traditional epidemiologic approach of evidence-based medicine owing to its high processing speed and ability to handle large volumes of data with heterogeneous variables (electronic health records, administrative data sets, wearable sensors, genomic and proteomic databanks, and social media) [18]. In fact, AI and ML have huge potential to build inferences and find patterns in vast volumes of patient histories, medical images, epidemiological statistics, and other particulars such as natural language data. For example, they can help doctors improve their diagnoses, forecast disease outbreaks, and customize treatments [19,20], provide better patient care [21], and predict the splicing activity of individual exons and chromatin marks from DNA sequences [22]. From a mental health perspective, the prevention of suicidal risk has recently been substantially studied [23-26].

Indeed, mental health care is also benefiting from the advancements in ML [27-29]. Classical ML with only mixed data (observations described by a mixture of numerical and categorical variables) is widely used, but language-based deficits are common symptoms of depression, bipolar disorder, autism spectrum disorder (ASD), personality disorder, and schizophrenia [30]. This implies that computational linguistics could have a great role in forming new insights into individuals' mental health and emotions.

Language in both spoken and written forms plays an important role in ML mental health applications. It is therefore essential to understand what natural language processing (NLP) is before discussing the joint applications of ML and NLP in mental health.

NLP

NLP is a subdiscipline of computer science that emerged in the 1960s. In 1967, the first published book on the subject, *Introduction to Computational Linguistics* [31], clearly considers language from a symbolic point of view: it describes techniques such as syntax parsing using dependency trees or Chomsky transformational grammars and statistical methods (word counting) are only hinted at. At that time, computing resources were sparse and had to be carefully managed; hence, a whole chapter of the book is dedicated to the storage of grammars in memory. The situation changed in the 1990s when personal computers became largely available and increasingly powerful. A new approach to NLP based on statistical methods emerged. The book by Manning and Schütze, *Foundations of Statistical Natural Language Process* [32], is a landmark of this evolution [32]. The 3 main sections of the book are dedicated to (1) methods at the word level (collocations, *n*-grams, and word sense disambiguation), (2) methods at the sentence level (morphosyntactic parsing using Markov models, and

probabilistic context-free grammars), and (3) clustering, classification, and information retrieval. Probabilistic context-free grammars are a typical example of the evolution of NLP methods: the symbolic approach by Chomsky—or at least a simplified version—is endowed with probabilities attached to productions, and the ambiguity of natural language is reflected in the coexistence of several syntax trees with different probabilities.

During the same period, symbolic methods evolved as well. The 1990s witnessed the emergence of the World Wide Web, the Semantic Web, and ontology engineering. First, the 2 research directions seemed contradictory. Knowledge representation was aimed at structuring knowledge in an exhaustively precise symbolic manner, whereas the statistical viewpoint considered language in the same way as physics considers natural phenomena: by analyzing them through various heteroclitic methods, identifying general laws by numerical indicators, and proving them using statistical methods. An example illustrating the latter is the distributional semantic hypothesis (originally stated in the paper by Harris titled, Distributional structure [33]) asserting that "Words occurring in the same contexts will tend to have related meanings." According to this hypothesis, one does not need to identify the precise meaning of a word, as a symbolic method would require, but simply to find the word's cooccurrences in a corpus and consider these as semantics of the word. A very popular method called latent semantic analysis (LSA) is based on the following: the matrix of occurrences of words in documents (contexts) is reduced so that the dimensions of the new matrix represent aggregates of words and aggregates of documents where each dimension is not interpretable per se, but when words or documents are represented as vectors in this new *latent* system of coordinates, the scalar product of vectors can be used as a semantic relatedness measure [34]. LSA is also an example of a typical ML method, with a learning phase (when the frequencies of words in the corpus are counted and the word or document matrix is reduced) to perform a specific task (evaluating the similarity between documents).

Since the 2000s and 2010s, a new evolution has occurred in NLP with the emergence of convolutional, recurrent, and recursive neural networks (NNs) [35]. By using large corpora and sophisticated learning approaches, these methods provide good performance in tasks of statistical nature, such as text classification or sentiment mining. In the past 3 years, they have been much more frequently used for learning higher syntactic or semantic structures (syntax graphs or concept mining, respectively).

In the future, hybrid methods may be used more frequently, which combine symbolic and statistical approaches. The presence of ML methods in NLP systems is a trend that will undoubtedly remain integral to contemporary methods through the foreseeable future.

Applications of ML and NLP to Mental Health

Applications of ML and NLP to mental health can be classified according to the following axes:

- The corpus: as one of the topics is NLP, the corpus necessarily has a textual component. The most common corpora are records or reports (electronic health records [EHRs], Psychological Evaluation Reports, and Coroner Reports), social media (Reddit, Twitter, etc), or transcribed patient interviews.
- Corpus processing: depending on the nature of the corpus, one can either extract medical terms and match them with unified medical language system (UMLS) concept unique identifiers (CUIs) or process blocks of text in natural language and perform specific searches (eg, to detect terms related to suicide).
- Classification methods: many ML techniques are used, such as decision trees, support vector machines, conditional random fields, random forests, and NNs.
- Goal: the goal is usually to validate a hypothesis or to study the behavior of a given population of patients.

Corpora can be of a very large volume. For instance, Sinnenberg et al [36] published a systematic review about Twitter as a tool for health research that included 137 different studies and analyzed over 5 billion tweets using ML; Castro et al [37] have processed 4.2 million EHRs spanning a period of over 20 years. Corpora can also be small, as demonstrated in the study conducted by Carson et al [38], who treated 73 respondents' unstructured clinical notes, or in the study by Bedi et al [39], in which only 34 participants' 1-hour-long narrative interviews were considered. Sometimes, corpora are created specifically for a project. For example, in a study by Roy et al [40], volunteers had written 182 abusive texts, annotated by researchers and abuse victims, and these texts were then analyzed and provided a model for detecting abusive texts.

Extraction of the UMLS CUIs is mainly applied to EHRs because the latter are semistructured and constitute a special document type. The specificities of this document type are reflected in its structure, the syntax of text, and, most importantly, the vocabulary used. The extraction of medical terms is achieved through information extraction algorithms and matching these terms with UMLS CUIs is performed through knowledge representation methods. Once these concepts have been extracted from an EHR, the latter is represented by the former and concepts become features used for classification.

On corpora other than EHRs, rather than extracting the UMLS CUIs, more general NLP methods are applied to textual data to obtain features that are then classified by ML algorithms. These NLP methods are often frequency counts of words or *n*-grams in a specific set, which can be manually curated or obtained out of a corpus. In other cases, methods such as LSA or latent Dirichlet allocation (LDA) are used for topic detection. The initial set of words can be explicit. For example, Doan et al [41] collected tweets containing the hashtags #relax and #stressed and classified them by theme and location. In other cases, calculations are performed at a higher level and words involved in the process are not explicitly known. For example, Luo et al [42] attempted to characterize autism by analyzing textual descriptions of closely related individuals written by patients or members of a control group. Nevertheless, most NLP applications in mental health rely on words (using the bag-of-words method, that is, ignoring word order and keeping

only their frequencies). Some take word order into account in a limited way (by using *n*-grams, ie, contiguous sequences of words of length *n*), but very few take syntax into account by the use of dependency trees [18,43,44]. With respect to their applications, it should be noted that ML and NLP tools are invaluable in alleviating data issues such as data overflow in modern medicine. Forsting et al [45] acknowledge that ML and NLP techniques can be useful for optimism bias (eg, the difference between a person's expectation and the actual outcome or the concept that a clinician may think that his or her patient's problem falls solely into a specific discipline in which the physician works) because the machine has a generalist approach unlike the specialist clinician. Within the last two decades, these techniques have emerged in mental health, following the success of social media to act as an informative source of data [46].

In addition, NLP is essential in psychiatry because language-based deficits are common symptoms of depression, behavioral disorder, ASD, personality disorder, and schizophrenia [30]. It can provide insight into individuals' mental health and emotions, their use of narrative, subjective, and structured speech styles, and their lifestyle, specifically their educational level, socioeconomic status, living conditions, and cultural background [47], all of which are routine in mental status examinations.

Using ML in general and NLP methods in particular, one can create semiautomated systems (operating under human supervision) aiming to improve the specificity of diagnosis, knowledge of psychophysiology, speed of diagnosis, and more accurate estimations of disease severity [48]. Through analyses of Twitter posts, O'Dea et al [49] identified the importance of creating real-time campaigns to increase help-seeking behaviors and reduce the stigma attached to mental health. Moreover, automated programs can be more cost-effective and time-efficient than their traditional counterparts. Ly et al [50] proposed using interventions based on an automated self-help system as a way to make mental health promotion tools more widely accessible. In addition, Lucas et al [51] demonstrated through a clinical trial that when people believed they were interacting with a computer rather than an actual clinician, they reported less fear of self-disclosure, reported reduced impression management behaviors, experienced more ease in expressing the severity of their emotions, and were rated by observers as more willing to disclose. However, these findings may not be generalizable, as they were potentially biased by their sample selections and/or system design itself.

Although ML and NLP provide new tools and strategies for psychiatric research and practice [52], it should be kept in mind that their use frequently raises ethical and legal concerns over consent to personal data use and data anonymization. Similarly, studies using AI for predictive analyses are challenging the balance between beneficence and respect for patients' autonomy. McKernan et al [53] suggest that efforts be made to communicate AI methods to obtain free and informed consent from patients. Moreover, prospective studies should be conducted to evaluate the use of AI tools [53].

The primary aim of this systematic review is to summarize and characterize studies that used ML and NLP techniques for mental health in methodological and technical terms. Hence, the secondary aim is to consider the potential use of these methods in mental health clinical practice, such as the contributions that they may offer in areas of diagnosis and prognosis, the establishment of risk factors, impacts of psychotherapy, treatment adherence, and side effects.

Methods

This systematic review is grounded in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analysis) guidelines [54]. Searches were carried out as specified by the standard protocol for PROSPERO (Prospective Register of Systematic Reviews; registration number CRD42019107376).

Literature Search Strategy

A systematic, computerized literature search was conducted using 4 databases: PubMed (via MEDLINE), Scopus, ScienceDirect, and PsycINFO. Each database was explored from August 21, 2018, through February 1, 2020, with no publication date limit. The search was carried out using the following keywords: "natural language processing" AND "machine learning" AND ("psychiatry" OR "mental health" OR "mental disorder"). The same search was performed on the element (data mining) instead of (machine learning). When the full text was not available, the abstract was used to extract the necessary information to avoid selection bias. Case studies, conference papers, and reviews were excluded.

Study Selection and Eligibility Criteria

After removing duplicates, 2 collaborators independently screened all titles and abstracts that were relevant to this systematic review. A third reviewer was consulted when disagreement arose between the first 2 collaborators. The process is depicted in Multimedia Appendix 1. Only studies available in English were selected. We deliberately excluded studies about the anonymization process to focus on the articles investigating the clinical use of ML and NLP in psychiatry (eg, contribution to diagnosis, prognosis, establishment of risk factors, impact of psychotherapy, treatment adherence, and side effect). No limitations on publication dates were imposed. A total of 58 articles were included in the review.

Included Studies

All studies were thoroughly screened, and their main ideas are summarized in individual tables (Multimedia Appendix 2 [37-41,43,47,48,55-104]). These tables provide information on qualitative and quantitative features: authors, year of publication, precise topic of mental health (eg, autism, psychotic spectrum disorder, etc), population characteristics, and types and volume of recorded data. The second part of these tables summarizes the objectives, methods, and results.

Results

Study Selection

The database search resulted in 222 studies identified using the (machine learning) keyword and 105 studies using the (data mining) keyword. After merging them, 238 unique studies were considered for review, based on the title and abstract. A total of 84 papers were excluded because (1) they were not about psychiatry or mental health (52 cases), (2) they were not written in English (1 case), and (3) the keywords (machine learning), (natural language processing), or (data mining) did not appear in the title or abstract (8 cases). As a second filter, 33 studies about data anonymization were excluded. Furthermore, 7 studies were excluded because ML or NLP were not their main subject but were only quoted as background information. In addition, 96 papers were excluded because they were reviews, case studies, or conference papers. Finally, 58 articles were included in this review.

Topics and Population

Topics are heterogeneous. The most frequently mentioned topics are depression and suicide with 17 studies [38,55,57,60-62,77-79,82,83,87,88,91,92,99,104]. Other psychiatric diagnoses were addiction to alcohol or illicit drugs (6 cases) [43,65,66,75,84,86]; posttraumatic stress disorder (PTSD; 3 cases) [47,63,64]; neurodevelopmental disorders (3 cases) [42,58,93]; psychotic spectrum disorders, including schizophrenia (3 cases) [39,95,100]; anxiety (2 cases) [41,98]; personality disorder (1 case) [85]; eating disorders (2 cases) [89,96]; and bipolar disorder (2 cases) [37,102]. A total of 3 studies were on violence and cyber harassment [40,80,94]. Treatment issues such as adherence or misuse are also depicted (6 cases) [56,72,74,81,101,103]. Only 1 study on mechanical restraints [90] and 1 on cognitive troubles [97] were found. A total of 8 studies were transnosographic [59,67-71,73,76]: 6 met the CEGS N-GRID 2016 Center of Excellence in Genomic Science Neuropsychiatric-Genome-Scale and Research Domain Criteria (RDoC) Individualized Domains 2016 Shared Task in Clinical NLP criteria, which will be developed further in our results.

In total, 3 distinct categories of population were found:

1. Patients whose EHRs were available in science-based research databases such as the Partners HealthCare electronic medical record (EMR), a collection of data from patients at Massachusetts General Hospital and Brigham and Women's Hospital [55,56]. These records extended beyond psychiatric records and included other medical records as well.
2. Patients seen in emergency or psychiatry departments who had additional clinical characteristics in their records (eg, clinical observation, laboratory tests, diagnostic and therapeutic interventions, typed specialists' notes).
3. Social media networks (Facebook, Twitter, and Instagram): The authors of these studies have selected specific hashtags such as #stress or #depression and have screened a multitude of public messages using a streaming platform.

Objectives

In total, 5 main categories of objectives were found: to extract clinical symptoms, to classify severity of illnesses, to compare different therapies, to provide psychopathological clues in mental health, and to challenge the current nosography.

The principal objectives of these studies were to extract and record clinical symptoms, establish a diagnosis, or monitor changes over time. A total of 2 studies targeted automated epidemiological monitoring: Metzger et al [57] provided a method of detecting suicide attempts from EHRs and Leroy et al [58] achieved automatic extraction of criteria for ASD from EHRs with an accuracy of 76%. The latter study stated that an increasing prevalence of given symptoms (nonverbal behavior, social and emotional reciprocity, and adherence to routine disabilities) occurred from 2000 through 2010. Data extraction was also used for diagnosis: He et al [47] diagnosed PTSD with an accuracy of 82% after analyzing free texts written by trauma survivors.

In addition to extraction, an important aim was to measure the severity of psychiatric disorders in psychological evaluation record corpora. Goodwin et al [59] classified symptoms of patients with psychosis into 4 different levels of severity (absent, mild, moderate, and severe) using statistical analyses. Fernandes et al [60] studied EHRs from a cohort of individuals with a history of suicide attempts and a cohort of individuals with a history of suicidal ideation only. Their algorithm of detecting suicidal ideation or suicide attempts had a sensitivity of 98.2% and a positive predictive value of 82.8% [57]. Other studies found that ML and NLP techniques performed well, although they were not necessarily better than a practitioner's ability to predict the clinical risk of suicide in their patients [61,62]; thus, the authors proposed statistical NLP approaches to be used in collaboration with clinical practice.

ML and NLP methods are also used to measure and compare the effectiveness of different types of psychotherapy [63,64]. Tanana et al [43] investigated 2 statistical NLP techniques to code motivational interviewing sessions. Motivational interviewing is a psychotherapy method used for substance use disorders and other behavioral problems to strengthen personal motivation for change [105]. Motivational interviews can be manually coded to assess therapy adherence and gather feedback for subsequent sessions. The authors found that the discrete sentence feature model (a sentence classifier based on *n*-gram models) had accuracy similar to the manual coding of therapeutic sessions. Maguen et al [63] used statistical NLP techniques to distinguish evidence-based psychotherapy, including cognitive processing therapy and prolonged exposure notes from unstructured psychotherapy notes for a population of veterans with PTSD. They found that almost 20% of veterans observed an improvement in their symptoms after one or more sessions of evidence-based psychotherapy.

Another objective was to provide psychopathological clues for understanding mental health disorders by analyzing language features. This objective sometimes involves the processing of previously unexplored data, such as chat groups or social networks. The following are some examples of studies that pursue this objective: Baggott et al [65] found that MDMA

(3,4-méthylènedioxy-N-méthylamphétamine; Ecstasy) altered individuals' speech patterns more frequently than the placebo and led to an increase in both positive and negative social and sexual language use (others, public, camaraderie, and outgoing). Chary et al [66] analyzed posts on Lycaeum, a popular web forum known for being one of the most frequently cited platforms with respect to drug use. They discovered new combinations of drugs that were not mentioned in the medical literature. Luo et al [42] differentiated the social interactions between adults with ASD and healthy adults. They confirmed the hypothesis regarding differences in language and social interactions in adults with ASD: typical participants had more connected semantic links than the ASD group and the words with the largest number of connections were different between the 2 groups. Doan et al [41] noticed that American Twitter users are more likely to express their source of stress on Twitter than in their day-to-day experiences. The main causes of stress that emerged from the Twitter data were education, work, and social relationships. They also found that individuals' expressions of stress and relaxation differed based on the city of residence (Los Angeles, New York, San Diego, and San Francisco). Moreover, Mowery et al [106] revealed that less than 2% of the tweet corpus (a corpus of 9300 annotated tweets containing depression-related keywords) included more than one depression-related reference, suggesting that there may be different forms of expression when it comes to depression.

Finally, AI in mental health research challenges the current practice and nosography. In 2010, Insel et al [107] initiated a project called the RDoC, a research framework for mental health disorders that aims to constitute an alternative to the DSM (Diagnostic and Statistical Manual of Mental Disorders). The former includes data on genetics and neuroscience in its classification of mental health disorders, whereas the latter is solely based on clinical data [107]. The RDoC is a matrix in which the columns and rows represent constructs (genes, molecules, cells, circuits, physiology, behaviors, self-reports, and paradigms) and subconstructs of each of the following 6 domains: negative valence, positive valence, cognitive systems, systems for social processes, arousal or regulatory systems, and sensorimotor systems. Pro-RDoC practitioners argue that DSM syndromes have significant limitations when used as phenotypes for identifying biomarkers and specific genetic variants associated with mental illness [108]. A concrete application of this new system used statistical NLP methods to create a phenotypic homogenous cohort that allowed a better comparison [109]. In 2016, the CEGS N-GRID (Centers of Excellence in Genomic Science Neuropsychiatric-Genome-Scale and RDoC Individualized Domains) proposed 3 challenging tasks using NLP methods: (1) data anonymization, (2) predicting symptom severity in the positive valence domain from neuropsychiatric clinical records, and (3) novel data use case (eg, predicting the presence of common mental conditions in patients) [67]. This research on NLP and ML processing identified 6 articles [59,67-71] that met these challenge tasks, although only 1 study dealt with task 3 [67]. As mentioned earlier, studies on anonymization were excluded; thus, the RDoC framework links the neuro-biological basis of mental processes with phenotypical manifestations [110]. The CEGS N-GRID shared task provided usable data for investigating ML and NLP techniques, which could lead to new psychiatric nosology.

Type of Data Used

As can be seen in Table 1 (in which no hapaxes are displayed), the most frequent corpus type is that of EHRs (to which EMRs can be added). EHRs (and EMRs) are convenient data sources because of their heterogeneity: they combine structured, semistructured, and free data, and they often use a significantly controlled language containing medical terms that allow the extraction of CUIs (Methods Section). The second most frequent sources of data are clinical notes and clinical records, which share the convenient properties of EHRs or EMRs, but are not standardized in the same way.

Table 1. Corpus type.

Characteristics	Values
EHRs[a]	22.9508
ClinNotes	16.3934
ClinRecords	11.4754
Interviews	8.1967
Tweets	8.1967
Questionnaires	6.5574
Reddit	6.5574
Web	4.918
EMRs[b]	3.2787

[a]EHR: electronic health record.

[b]EMR: electronic medical record.

The data described earlier share an important property: the corpora are generated by practitioners and therefore can be used for medical term extraction with satisfactory results.

A different category of data is generated by the patients. This category can be divided into 2 subcategories: data generated with the help of practitioners (eg, interviews and questionnaires)

and data freely generated by patients on social media (tweets, posts on Reddit, web blogs, etc).

Interviews and (textual parts of) questionnaires are technically free text but practitioners still have some amount of control over the content, and the environment in which the data are collected influences the degree of informality of texts. For these reasons, traditional NLP methods can be applied to them with satisfactory results.

Data collected from social media, because of their high degree of informality, loose spelling and syntax, and use of abbreviations and emojis, can only be superficially processed by standard NLP methods. Typical examples are in studies by Doan et al [41] and Jackson et al [73], in which tweets were selected because they contained the hashtags #stress and #relax and their words were used in a bag-of-words without any further linguistic treatment [41] or tweets were selected based on the presence of terms denoting opioids [72]. Although the authors lemmatized tweet contents, the main feature of tweets taken into account was their geographical origin.

Methodology

Two phases of NLP projects were distinguished: (1) preprocessing, which consists of analyzing the data to obtain numeric or categorical features, and (2) classification.

Preprocessing

Table 2 (in which no hapaxes are displayed) represents the frequency of use of various preprocessing methods that can be of different natures. Some methods apply to words or word groups: *lemma (*lemmatization, ie, replacing a word by a base form such as the singular for nouns or the infinitive for verbs), *POS* (part of speech, ie, attaching to a word a label denoting its grammatical function), *cTAKES* or *CUIs* (mapping a word or a noun phrase to concept in an ontology, such as the UMLS, and therefore unambiguously defining its semantics), *tf-idf* (attaching to a word or a term a value representing its significance in characterizing a given document or class it belongs to), *embedding* (representing a word by a vector in a high-dimensional space), *named-entity recognition* (deciding whether a given word or noun phrase is a named entity), *LIWC* (Linguistic Inquiry and Word Count, a commercial tool advertised as being “based on solid science” providing various “social and psychological insights” of words). Other methods combine words into higher structures: *n-grams* (considering an *n*-gram, ie, a sequence of *n* subsequent words, as an entity and measuring the frequencies of these entities). Finally, other methods are applied to entire sentences, paragraphs, or documents: *SentiAna* (analyzing sentiments or emotions), *LDA* and *LSA* (calculating sets of topics, detecting the significance of each topic for a given document, and providing representative words for each topic). The most frequent preprocessing methods are the standard methods of NLP (lemmatization, part-of-speech tagging, *n*-grams, and tf-idf), and methods specific to medical texts such as CUI extraction (keywords cTAKES and CUIs in Table 2). The embedding method is related almost exclusively to NNs and therefore is relatively recent. Finally, the tail of the graph in Table 2 contains methods applied primarily to free texts such as topic detection, named-entity recognition, sentiment or emotion analysis.

Table 2. Preprocessing methods.

Characteristics	Values
lemma	16.3043
POS[a]	10.8696
cTAKES[b]	10.8696
ngrams	9.7826
tfidf	7.6087
embedding	6.5217
CUIs[c]	5.4348
LDA[d]	5.4348
SentiAna	5.4348
LIWC[e]	4.3478
NER[f]	4.3478
LSA[g]	3.2609

[a]POS: part of speech.

[b]cTAKES: clinical Text Analysis and Knowledge Extraction System.

[c]CUI: concept unique identifier.

[d]LDA: latent Dirichlet allocation.

[e]LIWC: Linguistic Inquiry and Word Count.

[f]NER: named-entity recognition

[g]LSA: latent semantic analysis.

Classification

Once the classification phase is reached, linguistic data are entirely converted into numeric data, and therefore, the choice of classifier depends on factors other than corpus type. Some of these factors include (1) the volume of data, (2) the type of classification (supervised vs unsupervised), (3) the explicability level, and (4) the platform used. In Table 3 (where hapaxes are not displayed), we have shown decision tree, association rules, and C4.5 (also a decision tree algorithm) that are *transparent* methods, that is, the user can follow the classification process in a step-by-step manner and understand the reason a given individual belongs to a particular class. They are not the most frequent classifiers, probably because explicability is not a major concern of most studies. Instead, the most frequently used classifiers such as support vector machine (SVM), LogiR (logistic regression), RF (random forest), and LinR (linear regression) are solid, fast legacy classifiers with small parameter sets and good performance. In the middle of Table 3 are NNs that belong to the deep learning tendency of ML: they are opposite to DT/AR/C4.5 when it comes to explicability and they rely heavily on certain parameters (type and geometry of NN, number of layers, size of layers, optimizer, learning rate, loss function, etc). The causes of the relatively low frequency of NNs in publications may be (1) the fact that they have been implemented in user-friendly frameworks (such as Theano or Keras) only recently, (2) the necessity to fine-tune a large number of parameters, and (3) the relatively high requirements in terms of memory, central processing unit, and graphical processing unit. This is likely to change in the near future.

Table 3. Classifier type.

Characteristics	Values
SVM[a]	22.6804
LogiR[b]	16.4948
RF[c]	11.3402
DT[d]	6.1856
NB[e]	6.1856
NN[f]	6.1856
LinR[g]	5.1546
K-Means	3.0928
AR[h]	2.0619
C4.5	2.0619

[a]SVM: support vector machine.
[b]LogiR: logistic regression.
[c]RF: random forest.
[d]DT: decision tree.
[e]NB: Naive Bayes.
[f]NN: neural network.
[g]LinR: linear regression.
[h]AR: association rules.

Platforms

As can be seen in Table 4 (hapaxes are not represented), the 2 most common platforms are Python and R. Python is a *universal* programming language, in the sense that it is not specific to a given domain: more than 120,000 packages allow the user to perform specialized tasks in any possible field. Furthermore, it is open-source and high-quality documentation abounds. R is also an open-source programming language and compiler, but contrary to Python, it is oriented toward statistics. Although many classifiers have been implemented efficiently both in Python and R, the domain of NLP is better represented in Python, in credit to packages such as NLTK (Natural Language ToolKit), spaCy, and Stanza. The third bar, titled *Unknown*, represents publications that do not mention the platform used. The fourth bar indicates the General Architecture for Text Engineering General Health platform, an open-source Java application that provides an environment for processing textual data in a user-friendly manner. The *Apache* bar gathers different tools distributed by the Apache Software Foundation. Stata is a commercial statistics software from College Station, Texas, first released in 1985. Weka is an open-source programming environment for ML.

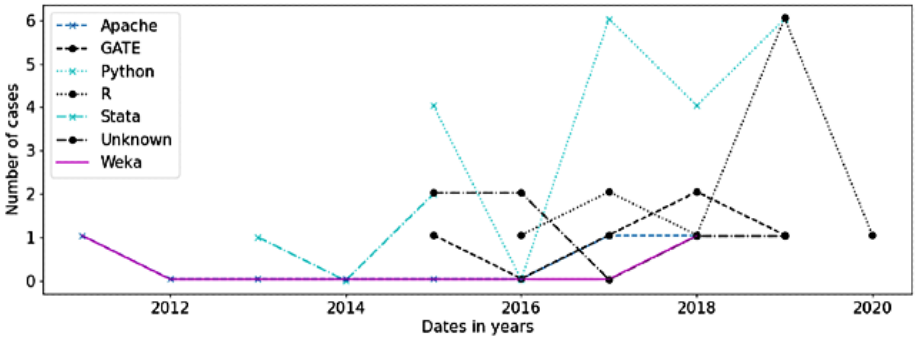
Figure 1 shows the use of platforms in chronological order. The use of Python and R started after 2015, while Stata, Weka, and Apache were already in use in 2011.

Table 4. Platforms.

Characteristics	Values
Python	34.4828
R	18.9655
Unknown	10.3448
GATE[a]	8.6207
Apache	5.1724
Stata	5.1724
Weka	3.4483

[a]GATE: General Architecture for Text Engineering General Health.

Figure 1. Platforms usage.

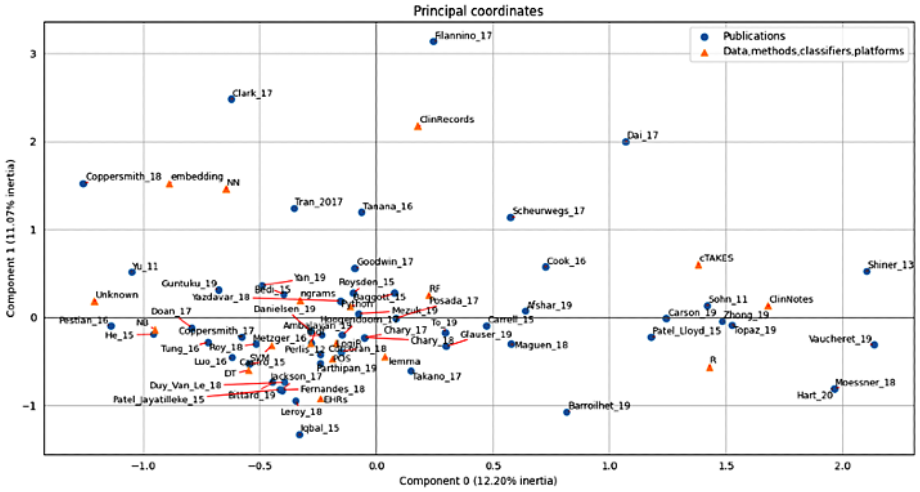


Correspondence Analysis of Data, Methods, Classifiers, Platforms, and Publications

The correspondence analysis is a dimension reduction technique that maps the data into a factorial space where each dimension is a combination of the initial variables. Figure 2 represents the principal coordinates of the publications and the various entities considered in their study.

On the right, a cluster of publications is surrounded by data type *ClinNotes*, method *cTAKES*, and platform *R*. In the upper left quadrant, some publications gather with method *embedding* and classifier *NN*. Toward the left of the diagram and close to the horizontal axis, publications with an *unknown* platform using the *NB* classifier are present along with a big cluster whose center includes *tf-idf*, *LogiR*, *SVM*, *Python*, and *n-grams*: the legacy methods, most used classifiers, and the most used platform.

Figure 2. Correspondence analysis.



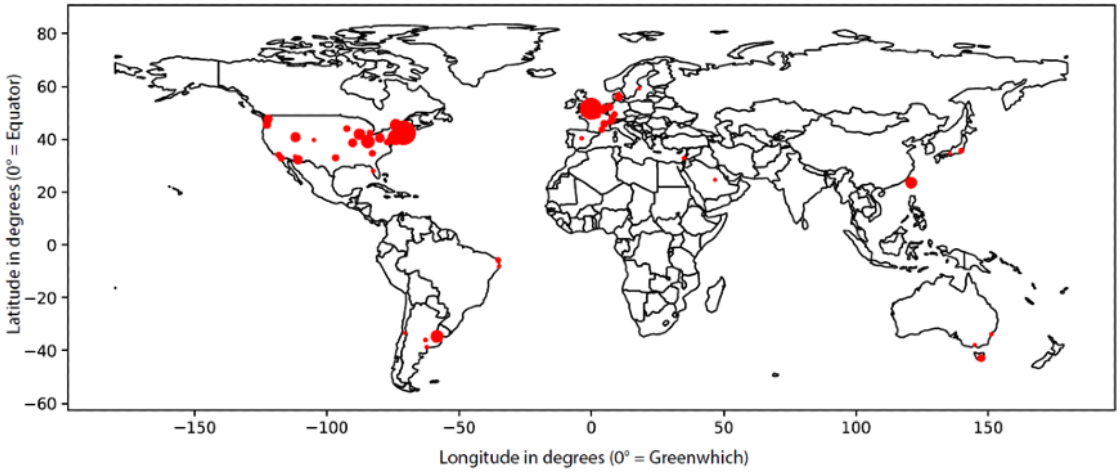
With regard to publications, *Filannino_17* is an obvious outlier because it has no method, classifier, or platform and because it describes a task and how this task has been treated by others. *Clark_17* is at the extreme upper left, as it uses NNs and k-means (the latter is not displayed because only entities appearing at least 5 times are included). *Coppersmith_18* also uses embeddings and NNs, whereas *Tran_17* (which is closer to the central cluster) uses both NNs and SVMs. On the right side, *Shiner_13* and *Vaucheret_19* use clinical notes and R, whereas *Hart_20* and *Moessner_18* use R and methods that

have not been taken into account in the calculation. In the bottom left, *Iqbal_15* uses EHRs in the General Architecture for Text Engineering General Health (which is not displayed). At the extreme left and close to the horizontal axis, *Pestian_16* and *Yu_11* use an unknown platform.

Geographical Distribution of Authors

In the map in Figure 3, the diameter of the red marks is proportional to a score calculated as follows: we added 1 unit for the geographical origin of the affiliation of each author of each paper. The cities with scores greater than 10 were Boston (54), London (44), New York (21), Cincinnati (15), Buenos Aires (13), Cambridge, Massachusetts (12), San Francisco (11), and Taiwan (11).

Figure 3. Geographical distribution of authors.



Citations and Cocitations

Figure 4 represents the citations of the papers in our list by other papers on the same list. The size of the nodes of a paper is proportional to the number of papers citing it. The colors of the nodes and edges represent communities. Each community has a central node: Perlis et al [55] are cited in 7 other papers, Jackson et al [73] are cited in 4 other papers, Carrell et al [74] and Afshar et al [75] are cited in 2 other papers, and Bedi et al [39] are cited in 2 other papers. In total, 22 papers are singletons: they are neither cited nor cite any other paper in our list.

Although mutual citations show influences between papers in our list, we can also measure the number of cocitations (ie, common references between 2 papers in the list). In Figure 5, the edges between papers indicate that they have at least 3 common references. The edge width is proportional to the number of references.

The edge of the greatest width is the one between the papers by Coppersmith et al [76] and Coppersmith et al [77], which is normal—the 2 papers share the same first author, have been released within less than a year, and have 26 common references.

The second case, in descending order of edge width, is between Shiner et al [64] and Maguen et al [63]. This is also normal—the first author of the former is also the last author of the latter and the latter is presented as an extension of the former: “In this study, our goal was to extend Shiner and colleagues’ work by applying automated coding to a large national pool of mental health treatment notes in order to identify the use of cognitive processing therapy and prolonged exposure.” The 2 papers share 14 common references.

The size of nodes in the graph is proportional to the degree. Zhong et al [78] have the highest degree: this paper has more than three common references with as many as eight other papers, in fact, with 8 references. The color of the nodes and edges corresponds to the connected components. There are 19 singleton nodes that share ≤2 references with every other paper of the list.

Figure 4. Graph of cocitations.

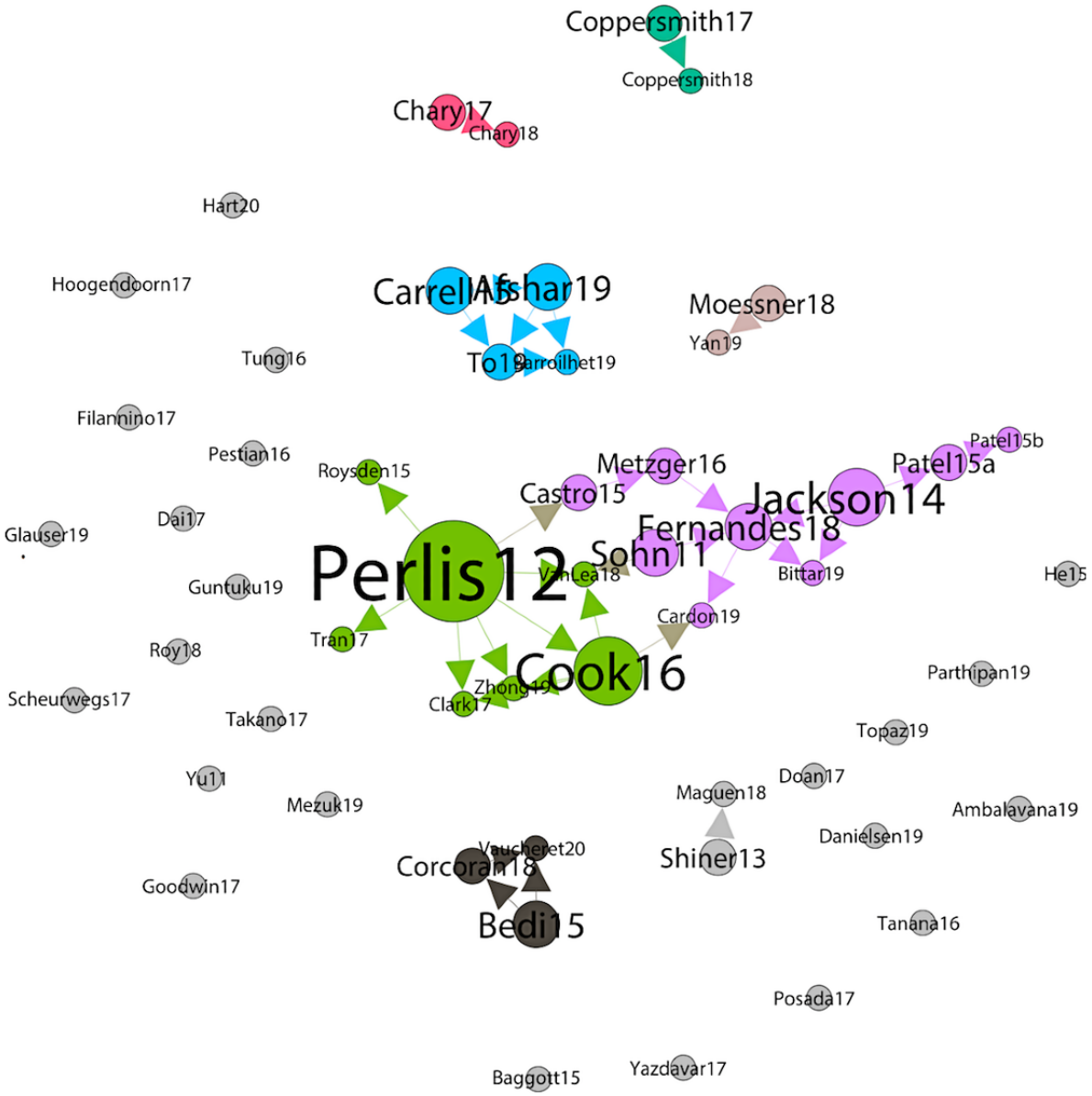
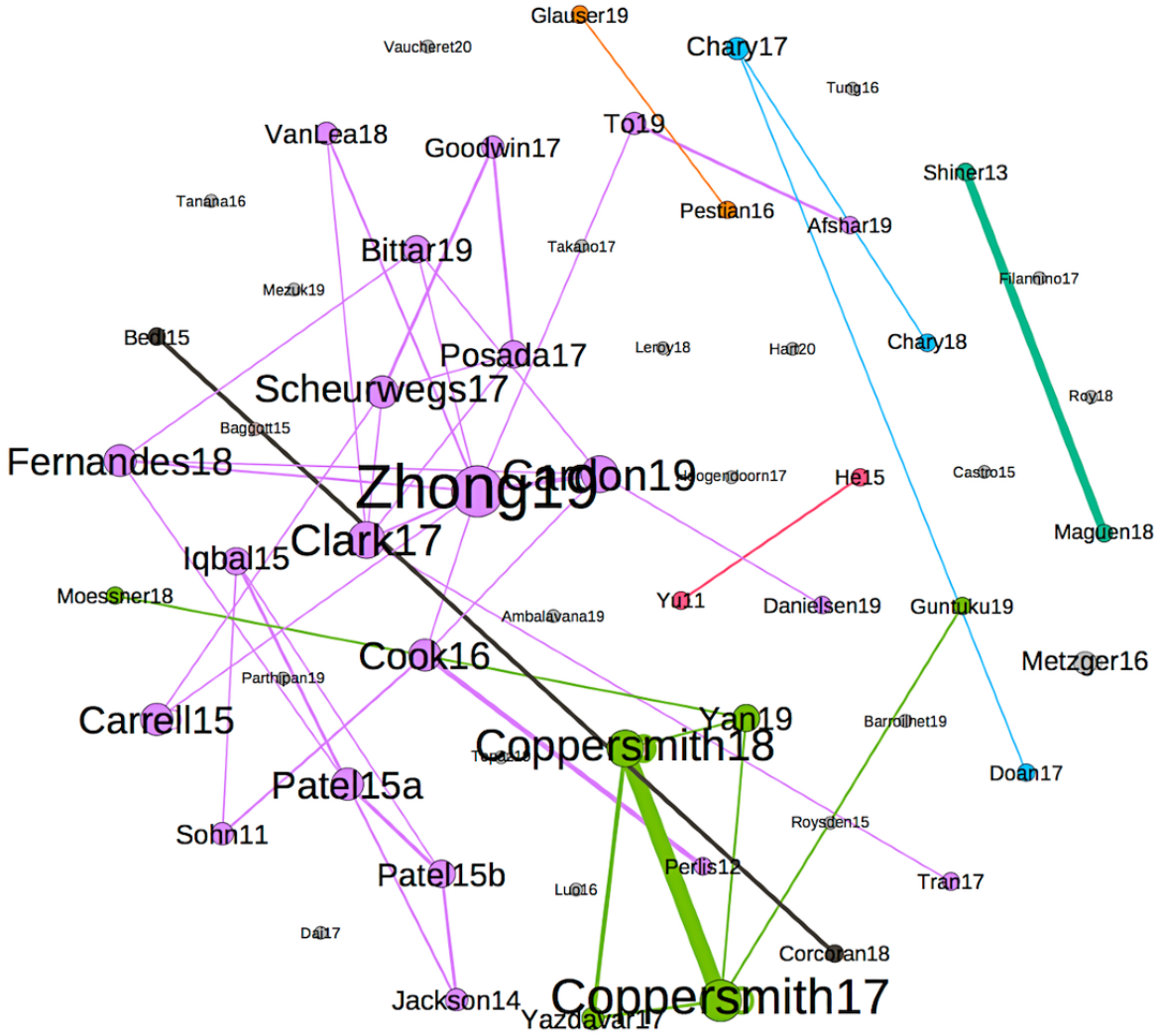


Figure 5. Graph of cocitations.



Discussion

Strengths and Limitations of the Review

This study reviews ML and NLP models in the field of mental health, which has been a highly topical issue in recent years. The methodology was elaborated to screen a maximum number of specific medical studies by expanding the research to 4 medical databases (PubMed, Scopus, ScienceDirect, and PsycINFO). Furthermore, the characterization of the selected studies has been done very precisely in a qualitative manner to simultaneously depict the populations, methods, data sources, and technical aspects.

The primary limitation of this study is the lack of quantitative comparisons between the selected studies. It is indeed not feasible to compare highly heterogeneous studies that do not share common research patterns. In addition, the selected works were not scored on their risk of bias. Despite this shortcoming, their limitations and strengths are outlined in the individual tables in Multimedia Appendix 2.

Methodological and Technical Limitations of the Selected Studies

ML and NLP methods may be considered as a new paradigm in medical research in which it becomes practical to analyze every possible, even unexpected, and innovative parameter of a topic to discern new clinical patterns. This new paradigm involves reconsidering the standard methodology, which consists of formulating a sound hypothesis, defining objectives, and collecting results to either uphold or reject the hypothesis. However, in practice, the selected studies tend to confirm clinical hypotheses based on fundamental clinical intuitions, namely language abnormalities in adults with ASD [42].

Other methodological limitations and potential bias sources have been noted. As stated in the Results section, one of the 3 main population categories is *social network or chat users* [40,41,66,77,79], whose members are predominantly young. Owing to this, Coppersmith et al [76,77] cautioned that these results may not be generalizable to other populations [77,106]. In addition, when Chary et al [66] focused on *Lycaeum users* and Coppersmith et al [76] mentioned *participants from a company*, the lack of precise information on the participants of a cohort was obvious. An exception to this is the group of OurDataHelps.org users [77] who volunteered to participate in

scientific research and filled out a questionnaire to provide information about themselves. Even when participants volunteer to provide personal information, there is a high likelihood that personality bias plays a role, especially in studies on suicide and depression.

Similarly, studies rarely consider cultural or ethnic differences within a sample [80]. For example, in a study on violent behavior, researchers should acknowledge that *spanking children for discipline purposes* is considered inappropriate in some cultures but appropriate in others. In some cases, language-specific features can improve the performance of NLP methods. For example, in the case of Takano et al [62], the distribution of morphemes is used to distinguish between specific and nonspecific memories in the Autobiographical Memory Test. As shown in the paper, among the most important distinctive factors are grammatical particles that are specific to the Japanese language, such as た/だ (past tense), ない (negation), は (topic marker), and で (place or method). In languages with different structures, the same method may be less efficient and other indicators may need to be investigated.

Is There an Advantage in Using ML and NLP for Mental Health Clinical Practice?

The hallmark ML principle is to simultaneously analyze large quantities of data; however, this sometimes leads researchers to the implicit assumption that *the more data they input, the more accurate will be the results*. ML and NL allow the analysis of large amounts of data and the comparison of broad groups and patients. For example, Roysden et al [56] screened administrative data and EHRs from a population of 12,759 patients; Maguen et al [63] compared over 8,168,330 clinical notes collected over 15 years; and Yazdavar et al [79] analyzed posts authored by 4000 Twitter users. At the same time, even though thousands of papers have been published using medical data, very few have made meaningful contributions to clinical practices [111].

Twitter and other social networks, with almost 3 billion users globally, have become significant sources of information for medical use and research [112]. Moreover, the analysis of social media–based platforms can generate valuable details about people's mental health and social or professional interactions. The alteration of daily habits is one of the core criteria for the diagnosis of a mental health disorder (in general, criterion B of DSM-5). A recent study by Fagherazzi and Ravaud [113] illustrates the idea that AI can be implemented in the so-called *digitosome* (data generated online and by digital technologies) that constitutes a powerful agent for detecting new digital markers and risk factors in medicine. By analyzing a global cohort of more than 85,000 tweets per week authored by people with diabetes, they were able to discuss different illness-related stress patterns of patients with type 1 or type 2 diabetes. By analyzing tweets, Mowery et al [106] found that there may be alternative ways in which people express depression. These findings indicate that there may be new ways for people to express mental illness.

From this perspective, different expressions of psychological distress (whether people are addressing health care professionals, relatives, or digital friend networks) could be accessible and useful to care providers. ML and NLP may be valuable in psychiatry for identifying people with clinical risks for depression, suicide attempts, anxiety, or even psychosis based on digital data or clinical notes.

Ethical Reflections

AI in psychiatry and more broadly in medicine raises ethical issues and requires prudence in its application. As mentioned earlier, ML and NLP techniques have valuable advantages in psychiatry for analyzing large amounts of data with high diagnostic and prognostic validity. These tools, which have been groundbreaking in medicine and psychiatry, should receive more attention for their promising results with regard to clinical practice and medical research. In addition, recent studies suggest that people are becoming more comfortable when speaking with a machine compared with a clinician: Lucas et al [51] state that in a clinical trial, people who (believed they) were interacting with a computer disclosed information more openly than people who thought that an individual was controlling the computer. Perhaps the machine is viewed as being more objective than a human and therefore reduces the fear of judgment from a practitioner. The introduction of a computer in medical practice as a new type of clinician leads to a profound change in the physician-patient relationship and promotes the idea of having a new clinical model involving a third party. The relationship is crucial to psychiatric clinical practice, and the use of data processing should be discussed. Sassolas [114] questioned this *technological psychiatry* as a practice that is likely to avoid what he called the "psychic privacy proximity." Technological psychiatry could generate an operative encounter whose unique purpose is to normalize the patient's symptoms and reduce the fear of disclosure.

In addition to improved relationships, the application of ML and NLP in psychiatry should be done with special precautions to avoid clinical abuse. This review includes 2 studies about the prediction of psychosis in patients at high risk of this disease. One even introduced a model of ML+NLP that had a 100% accuracy in predicting psychosis among the latter patient sample [39], which was better than a simple clinical evaluation. Nevertheless, these results should be treated with caution because of the small sample size and the lack of detail on the statistical techniques used. The risk of overfitting needs to be considered. Although further research should be continued to improve technical issues, ethics should be taken into account. Martinez-Martin et al [115] questioned whether it is ethical to use prognostic estimates from ML to treat psychosis, as it is not known whether variables are present in the local context (such as differences in psychiatric practice and social support) that would affect the model's validity. Moreover, when programming an ML algorithm, investigators can choose to strengthen the criteria they esteem to be more relevant, such as clinical criteria instead of socioeconomic factors. This could result in loss of opportunity for some patients when the automated machine analysis gives the illusion of greater objectivity. These adjustments should be done to respect the principle of equity.

In the case of predicting psychosis, the study involved only patients who consented to both psychiatric care and the completion of interviews. This was not the case in studies on

suicide prevention, where researchers tracked information on patients by using social media. This could be considered a violation of confidentiality. Should information from social media be used to identify symptoms? Applying AI in this context raises significant ethical concerns, particularly in balancing beneficence and respecting confidentiality [53]. ML and NLP can help identify people at clinical risk for depression or suicidal ideation, who most likely do not have access to mental health providers and/or a primary care doctor [61]; however, this reduces confidentiality protection and can lead to increased vulnerability in certain populations [21]. To obtain informed consent from patients and protect their privacy, McKernan et al [53] proposed some recommendations: patients should be informed that (1) algorithms can be imperfect or wrong; (2) algorithm data should be considered highly sensitive or confidential; (3) algorithm data might recommend actions that are not immediately apparent; and (4) algorithms might prompt unnecessary intervention from the provider. Therefore, psychiatrists should be trained in ML and NLP techniques and be able to explain to patients their main characteristics and why they may require certain recommendations. This last point underlines the need for an explainable AI that goes further than black box methods.

Finally, ML and NLP should not lead to disempowerment of psychiatrists or replace the clinician-patient pair. On the contrary, the combination of ML with NLP should be considered as a *tool* to support clinical practice and medical research.

Conclusions

In the past decade, the use of ML and NLP has become increasingly widespread in medicine and more specifically in psychiatry. Hence, this review aimed to summarize and characterize studies that used ML and NLP techniques for mental health in methodological and technical terms. The secondary aim was to consider the potential use of these methods in mental health clinical practice (eg, contribution to diagnosis, prognosis, establishment of risk factors, impact of psychotherapy, treatment adherence, and side effects).

Although the selected studies were heterogeneous in terms of topics and mental disorders, common features were found in terms of population categories (patients included in medical databases, patients presenting to the emergency room, and social media network users) and objectives (ie, symptom extraction, severity classification, comparison of therapies, findings of psychopathological clues, and challenges to the current nosography). The type-of-data-used analysis identified 2 major corpora: data collected by care providers (EHR, clinical notes, or EMR) and data from social media. Finally, the method analysis indicates that the authors privileged certain techniques. The standard methods of NLP (such as lemmatization, POS tagging, or n-grams) are most frequently used for preprocessing, in addition to CUI extraction dedicated to medical texts. The classification analysis specifies that classifiers with good performance (SVM, LogIR, and RF) are preferred to those with *transparent* functioning. The use of the universal programming language platforms such as Python and R is verified; Python turned out to be the most frequently and recently used. The correspondence analysis of data, methods, classifiers, platforms, and publications reveals a cluster of publications associating clinical notes data with cTAKES methods and the R-Python platform.

ML and NLP methods may sometimes be impressive with their huge amount of data screening and the multiple perspectives they offer. This has led some authors to consider it to be a new paradigm in mental health research. However, these processes tend to confirm clinical hypotheses rather than developing new information, and some results should be treated with caution (eg, results from social media users' cohorts or the impact of language-specific features on NLP methods performance). On the contrary, ML and NLP techniques provide information from unexplored data and on patients' daily habits that are usually inaccessible to care providers. It may be considered as an additional tool in every step of mental health care: diagnosis, prognosis, treatment efficacy, and monitoring. In this regard, ethical issues, such as predicting psychiatric troubles or implications in the physician-patient relationship, remain and should be discussed in a timely manner. Therefore, ML and NLP methods may offer multiple perspectives in mental health research, but they should be considered as a tool to support clinical practice.

Acknowledgments

This work has been supported by the French government's "Investissements d'Avenir" programme, which is managed by the Agence Nationale de la Recherche (ANR), under the reference ANR-18-RHUS-0014.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Preferred reporting items for systematic reviews (PRISMA) flow diagram.
[PNG File , 160 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Table summarizing the selected studies.
[PDF File (Adobe PDF File), 338 KB-Multimedia Appendix 2]

References

1. Domingos P. A few useful things to know about machine learning. Commun ACM 2012 Oct;55(10):78-87. [doi: 10.1145/2347736.2347755]
2. Dada EG, Bassi JS, Chiroma H, Abdulhamid SM, Adetunmbi AO, Ajibuwa OE. Machine learning for email spam filtering: review, approaches and open research problems. Heliyon 2019 Jun;5(6):e01802 [FREE Full text] [doi: 10.1016/j.heliyon.2019.e01802] [Medline: 31211254]
3. Lessmann S, Baesens B, Seow H, Thomas LC. Benchmarking state-of-the-art classification algorithms for credit scoring: an update of research. Eur J Oper Res 2015 Nov;247(1):124-136. [doi: 10.1016/j.ejor.2015.05.030]
4. Danuta Z, Jan M. Clustering Algorithms for Bank Customer Segmentation. In: 5th International Conference on Intelligent Systems Design and Applications. 2005 Presented at: ISDA'05; June 26-27, 2005; Warsaw. [doi: 10.1109/isda.2005.33]
5. Afreen M, Badugu S. Document clustering using different unsupervised learning approaches: A survey. In: Advances in Decision Sciences, Image Processing, Security and Computer Vision. Cham, UK: Springer; 2000.
6. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
7. Chen Q, Li H, Tang B, Wang X, Liu X, Liu Z, et al. An automatic system to identify heart disease risk factors in clinical texts over time. J Biomed Inform 2015 Dec;58 Suppl:S158-S163 [FREE Full text] [doi: 10.1016/j.jbi.2015.09.002] [Medline: 26362344]
8. Fraser KC, Meltzer JA, Graham NL, Leonard C, Hirst G, Black SE, et al. Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. Cortex 2014 Jun;55:43-60. [doi: 10.1016/j.cortex.2012.12.006] [Medline: 23332818]
9. Oliva J, Serrano JI, del Castillo MD, Iglesias A. A methodology for the characterization and diagnosis of cognitive impairments-application to specific language impairment. Artif Intell Med 2014 Jun;61(2):89-96. [doi: 10.1016/j.artmed.2014.04.002] [Medline: 24813116]
10. Mahabub A. A robust voting approach for diabetes prediction using traditional machine learning techniques. SN Appl Sci 2019 Nov 25;1(12):1. [doi: 10.1007/s42452-019-1759-7]
11. Sonark P, JayaMalini. K. Diabetes Prediction Using Different Machine Learning Approaches. In: 3rd International Conference on Computing Methodologies and Communication. 2019 Presented at: ICMMC'19; May 6-9, 2019; Erode, India. [doi: 10.1109/ICCMC.2019.8819841]
12. Sharma M, Singh G, Singh R. An advanced conceptual diagnostic healthcare framework for diabetes and cardiovascular disorders. ICST Transactions on Scalable Information Systems 2018 Jun 19;5(18):154828. [doi: 10.4108/eai.19-6-2018.154828]
13. Cruz J, Wishart D. Applications of Machine Learning in Cancer Prediction and Prognosis. Cancer Inform 2017 Oct 04;2:117693510600200-117693510600217 [FREE Full text] [doi: 10.1177/117693510600200030]
14. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015;13:8-17 [FREE Full text] [doi: 10.1016/j.csbj.2014.11.005] [Medline: 25750696]
15. Gauthier E, Brisson L, Lenca P, Ragusa S. Breast Cancer Risk Score: a Data Mining Approach to Improve Readability. In: The International Conference on Data Mining. 2011 Presented at: CDM'11; June 23-27, 2011; Las Vegas, United States p. 15-21.
16. Gauthier E, Brisson L, Lenca P, Clavel-Chapelon F, Ragusa S. Challenges to Building a Platform for a Breast Cancer Risk Score. In: Sixth International Conference on Research Challenges in Information Science. 2012 May Presented at: RCIS'12; June 1-7, 2012; Valencia, Spain.
17. Dhahri H, Al Maghayreh E, Mahmood A, Elkilani W, Faisal Nagi M. Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms. J Healthc Eng 2019;2019(2):4253641-4253643 [FREE Full text] [doi: 10.1155/2019/4253641] [Medline: 31814951]
18. Scott IA. Machine learning and evidence-based medicine. Ann Intern Med 2018 Jul 3;169(1):44-46. [doi: 10.7326/M18-0115] [Medline: 29710098]
19. Bughin J. Artificial intelligence, the next digital frontier? McKinsey Global Institute 2017 Jun:80.
20. David C, Sastry C, Martha I, Saif R, Jordan V. Ma- Chine Learning and Therapeutics 2.0: Avoiding Hype, Realizing Potential. McKinsey & Commpany. URL: https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/machine-learning-and-therapeutics-2-0-avoiding-hype-realizing-potential?cid=eml-app [accessed 2018-12-17]
21. Darcy AM, Louie AK, Roberts LW. Machine learning and the profession of medicine. J Am Med Assoc 2016 Feb 9;315(6):551-552. [doi: 10.1001/jama.2015.18421] [Medline: 26864406]
22. Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep learning for computational biology. Mol Syst Biol 2016 Jul 29;12(7):878 [FREE Full text] [doi: 10.15252/msb.20156651] [Medline: 27474269]
23. Lopez-Castroman J, Nogue E, Guillaume S, Picot MC, Courtet P. Clustering suicide attempters: impulsive-ambivalent, well-planned, or frequent. J Clin Psychiatry 2016 Jun;77(6):e711-e718. [doi: 10.4088/JCP.15m09882] [Medline: 27035768]

24. Oh J, Yun K, Hwang JH, Chae JH. Classification of suicide attempts through a machine learning algorithm based on multiple systemic psychiatric scales. Front Psychiatry 2017;8:192 [FREE Full text] [doi: 10.3389/fpsyt.2017.00192] [Medline: 29038651]
25. Berrouiguet S, Billot R, Larsen ME, Lopez-Castroman J, Jaussent I, Walter M, et al. An approach for data mining of electronic health record data for suicide risk management: database analysis for clinical decision support. JMIR Ment Health 2019 May 7;6(5):e9766 [FREE Full text] [doi: 10.2196/mental.9766] [Medline: 31066693]
26. Belsher BE, Smolenski DJ, Pruitt LD, Bush NE, Beech EH, Workman DE, et al. Prediction models for suicide attempts and deaths: a systematic review and simulation. JAMA Psychiatry 2019 Jun 1;76(6):642-651. [doi: 10.1001/jamapsychiatry.2019.0174] [Medline: 30865249]
27. Luxton DD. An introduction to artificial intelligence in behavioral and mental health care. In: Luxton DD, editor. Artificial Intelligence in Behavioral and Mental Health Care. San Diego CA: Academic Press; 2016:1-26.
28. Kaur P, Sharma M. Diagnosis of human psychological disorders using supervised learning and nature-inspired computing techniques: a meta-analysis. J Med Syst 2019 May 28;43(7):204. [doi: 10.1007/s10916-019-1341-2] [Medline: 31139933]
29. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med 2019 Jul;49(9):1426-1448. [doi: 10.1017/S0033291719000151] [Medline: 30744717]
30. Cohen AS, Mitchell KR, Elvevåg B. What do we really know about blunted vocal affect and alogia? A meta-analysis of objective assessments. Schizophr Res 2014 Nov;159(2-3):533-538 [FREE Full text] [doi: 10.1016/j.schres.2014.09.013] [Medline: 25261880]
31. Hays DG. Introduction to Computational Linguistics, Mathematical Linguistics and Automatic Language Processing. Cambridge: American Elsevier Publishing Co; 1967.
32. Manning CD, Schutze H. Foundations of Statistical Natural Language Processing. Cambridge: MIT Press Ltd; 1999.
33. Harris ZS. Distributional Structure. WORD 2015 Dec 4;10(2-3):146-162. [doi: 10.1080/00437956.1954.11659520]
34. Landauer TK, Foltz PW, Laham D. An introduction to latent semantic analysis. Discourse Processes 1998 Jan;25(2-3):259-284. [doi: 10.1080/01638539809545028]
35. Chollet F. Deep Learning with Python. Shelter Island: Manning Publications; 2017.
36. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
37. Castro VM, Minnier J, Murphy SN, Kohane I, Churchill SE, Gainer G, International Cohort Collection for Bipolar Disorder Consortium. Validation of electronic health record phenotyping of bipolar disorder cases and controls. Am J Psychiatry 2015 Apr;172(4):363-372 [FREE Full text] [doi: 10.1176/appi.ajp.2014.14030423] [Medline: 25827034]
38. Carson NJ. Identification of suicidal behavior among psychiatrically hospitalized adolescents using natural language processing and machine learning of electronic health records. PLoS One 2019 Feb 19;14(2):01-14. [doi: 10.1371/journal.pone.0211116] [Medline: 30779800]
39. Bedi G, Carrillo F, Cecchi GA, Fernandez Slezak D, Sigman M, Mota NB, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr 2015 Aug 26;1:15030. [doi: 10.1038/npjschz.2015.30] [Medline: 27336038]
40. Roy T, McClendon J, Hodges L. Analyzing Abusive Text Messages to Detect Digital Dating Abuse. In: IEEE International Conference on Healthcare Informatics. 2018 Presented at: IEEE'18; March 2-3, 2018; New-York. [doi: 10.1109/ICHI.2018.00039]
41. Doan S, Ritchart A, Perry N, Chaparro JD, Conway M. How do you #relax when you're #stressed? A content analysis and infodemiology study of stress-related tweets. JMIR Public Health Surveill 2017 Jun 13;3(2):e35 [FREE Full text] [doi: 10.2196/publichealth.5939] [Medline: 28611016]
42. Luo SX, Shinall JA, Peterson BS, Gerber AJ. Semantic mapping reveals distinct patterns in descriptions of social relations in adults with autism spectrum disorder. Autism Res 2016 Aug;9(8):846-853. [doi: 10.1002/aur.1581] [Medline: 26613541]
43. Tanana M, Hallgren KA, Imel ZE, Atkins DC, Srikumar V. A comparison of natural language processing methods for automated coding of motivational interviewing. J Subst Abuse Treat 2016 Jun;65:43-50 [FREE Full text] [doi: 10.1016/j.jsat.2016.01.006] [Medline: 26944234]
44. Haralambous Y, Lemey C, Lenca P, Billot R, Kim-Dufor DH. Using NLP Methods to Predict Psychiatric Comorbidities. In: the RaPID-3: Resources and ProcessIng of linguistic, para-linguistic and extra-linguistic Data from people with various forms of cognitive, psychiatric and/or developmental impairments Workshop. 2020 Presented at: RAPID'20; May 2-6, 2020; Marseille.
45. Forsting M. Machine Learning Will Change Medicine. J Nucl Med 2017 Mar;58(3):357-358 [FREE Full text] [doi: 10.2967/jnumed.117.190397] [Medline: 28153954]
46. Sarasohn-Kahn J. The wisdom of patients: health care meets online social media. Technical report, California HealthCare Foundation 2008 [FREE Full text]
47. He Q, Veldkamp BP, Glas CA, de Vries T. Automated Assessment of Patients' Self-Narratives for Posttraumatic Stress Disorder Screening Using Natural Language Processing and Text Mining. Assessment 2017 Mar;24(2):157-172. [doi: 10.1177/1073191115602551] [Medline: 26358713]

48. Alanazi HO, Abdullah AH, Qureshi KN. A Critical Review for Developing Accurate and Dynamic Predictive Models Using Machine Learning Methods in Medicine and Health Care. J Med Syst 2017 Apr;41(4):69. [doi: 10.1007/s10916-017-0715-6] [Medline: 28285459]
49. O'Dea B, Achilles MR, Larsen ME, Batterham PJ, Calear AL, Christensen H. The rate of reply and nature of responses to suicide-related posts on Twitter. Internet Interv 2018 Sep;13:105-107 [FREE Full text] [doi: 10.1016/j.invent.2018.07.004] [Medline: 30206524]
50. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot rct using mixed methods. Internet Interv 2017 Oct 10;10:39-46. [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
51. Lucas GM, Gratch J, King A, Morency LP. It's only a computer: Virtual humans increase willingness to disclose. Computers in Human Behavior 2014 Aug;37:94-100. [doi: 10.1016/j.chb.2014.04.043]
52. Wongkoblap A, Vadillo MA, Curcin V. Researching mental health disorders in the era of social media: systematic review. J Med Internet Res 2017 Jun;19(6):e228. [doi: 10.2196/jmir.7215] [Medline: 28663166]
53. McKernan LC, Clayton EW, Walsh CG. Protecting life while preserving libertythical recommendations for suicide prevention with artificial intelligence. Front Psychiatry 2018;9:650. [doi: 10.3389/fpsyt.2018.00650] [Medline: 30559686]
54. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic reviewmeta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015 Jan 2;350:g7647. [doi: 10.1136/bmj.g7647] [Medline: 25555855]
55. Perlis RH, Iosifescu DV, Castro VM, Murphy SN, Gainer VS, Minnier J, et al. Using electronic medical records to enable large-scale studies in psychiatry: Treatment resistant depression as a model. Psychological Medicine 2012 Jan;42(1):41-50. [doi: 10.1017/S0033291711000997] [Medline: 21682950]
56. Roysden N, Wright A. Predicting health care utilization after behavioral health referral using natural language processing and machine learning. 2015 Presented at: AMIA Annu Symp Proc; 2015; San Francisco p. 2063-2072 URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4765610/pdf/2246862.pdf
57. Metzger MH, Tvardik N, Gicquel Q, Bouvry C, Poulet E, Potinet-Pagliaroli V. Use of emergency department electronic medical records for automated epidemiological surveillance of suicide attempts: a French pilot study. Int J Methods Psychiatr Res 2017 Jun;26(2):e1522. [doi: 10.1002/mpr.1522] [Medline: 27634457]
58. Leroy G, Gu Y, Pettygrove S, Galindo MK, Arora A, Kurzius-Spencer M. Automated extraction of diagnostic criteria from electronic health records for autism spectrum disorders: Development, evaluation, and application. J Med Internet Res 2018 Nov 7;20(11):e10497. [doi: 10.2196/10497] [Medline: 30404767]
59. Goodwin TR, Maldonado R, Harabagiu SM. Automatic recognition of symptom severity from psychiatric evaluation records. J Biomed Inform 2017 Nov:71-74. [doi: 10.1016/j.jbi.2017.05.020] [Medline: 28576748]
60. Fernandes AC, Velupillai S, Sanyal J, Stewart R, Chandran D. Identifying suicide ideation and suicidal attempts in a psychiatric clinical research database using natural language processing. Sci Rep 2018 May 09;8(1):7426. [doi: 10.1038/s41598-018-25773-2]
61. Cook BL, Progovac AM, Chen P, Mullin B, Hou S, Baca-Garcia E. Novel use of natural language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid. Comput Math Methods Med 2016:8708434. [doi: 10.1155/2016/8708434] [Medline: 27752278]
62. Pestian JP, Grupp-Phelan J, Bretonnel Cohen K, Meyers G, Richey LA, Matykiewicz P, et al. A controlled trial using natural language processing to examine the language of suicidal adolescents in the emergency department. Suicide Life Threat Behav 2016 Apr;46(2):154-159. [doi: 10.1111/sltb.12180] [Medline: 26252868]
63. Maguen S, Madden E, Patterson OV, DuVall SL, Goldstein LA, Burkman K, et al. Measuring use of evidence based psychotherapy for post-traumatic stress disorder in a large national healthcare system. Adm Policy Ment Health 2018 Jul;45(4):519-529. [doi: 10.1007/s10488-018-0850-5] [Medline: 29450781]
64. Shiner B, D'Avolio LW, Nguyen TM, Zayed MH, Young-Xu Y, Desai RA, et al. Measuring use of evidence based psychotherapy for posttraumatic stress disorder. Adm Policy Ment Health 2013 Jul;40(4):311-8. [doi: 10.1007/s10488-012-0421-0] [Medline: 22535469]
65. Baggott MJ, Kirkpatrick MG, Bedi G, de Wit H. Intimate insight: MDMA changes how people talk about significant others. J Psychopharmacol 2015 Jun;29(6):669-677. [doi: 10.1177/0269881115581962]
66. Chary M, Yi D, Manini AF. Candyflipping and Other Combinations: Identifying Drug-Drug Combinations from an Online Forum. Front Psychiatry 2018;9:135 [FREE Full text] [doi: 10.3389/fpsyt.2018.00135] [Medline: 29760666]
67. Tran T, Kavuluru R. Predicting mental conditions based on «history of present illness» in psychiatric notes with deep neural networks. J Biomed Inform 2017 Nov;75:138-148. [doi: 10.1016/j.jbi.2017.06.010] [Medline: 28606869]
68. Scheurwegs E, Sushil M, Tulkens S, Daelemans W, Luyckx K. Counting trees in random forests: Predicting symptom severity in psychiatric intake reports. J Biomed Inform 2017 Nov;75:112-119. [doi: 10.1016/j.jbi.2017.06.007] [Medline: 28602906]
69. Posada JD, Barda AJ, Shi L, Xue D, Ruiz V, Kuan PH, et al. Predictive modeling for classification of positive valence system symptom severity from initial psychiatric evaluation records. J Biomed Inform 2017;75:94-104. [doi: 10.1016/j.jbi.2017.05.019] [Medline: 28571784]

70. Clark C, Wellner B, Davis R, Aberdeen J, Hirschman L. Automatic classification of RDoC positive valence severity with a neural network. Journal of Biomedical Informatics 2017 Nov;75:120-128. [doi: 10.1016/j.jbi.2017.07.005] [Medline: 28694118]
71. Filannino M, Stubbs A, Uzuner O. Symptom severity prediction from neuropsychiatric clinical records: Overview of 2016 CEGS N-GRID shared tasks Track 2. Journal of Biomedical Informatics 2017 Nov;75:62-70. [doi: 10.1016/j.jbi.2017.04.017] [Medline: 28455151]
72. Chary M, Genes N, Giraud-Carrier C, Hanson C, Nelson LS, Manini AF. Epidemiology from Tweets: Estimating Misuse of Prescription Opioids in the USA from Social Media. J Med Toxicol 2017;13(4):278-286. [doi: 10.1007/s13181-017-0625-5] [Medline: 28831738]
73. Jackson RG, Ball M, Patel R, Hayes RD, Dobson RJB, Stewart R. Texthunter- a user friendly tool for extracting generic concepts from free text in clinical research. 2014 Nov Presented at: AMIA Annu Symp Proc; 2014; Washington DC p. 729-738 URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4420012/pdf/1985470.pdf
74. Carrell DS, Cronkite D, Palmer RE, Saunders K, Gross DE, Masters ET, et al. Using natural language processing to identify problem usage of prescription opioids. Int J Med Inform 2015;84(12):1057-1064. [doi: 10.1016/j.ijmedinf.2015.09.002] [Medline: 26456569]
75. Afshar M, Philips A, Karnik N, Mueller J, To D, Gonzalez R, et al. Natural language processing and machine learning to identify alcohol misuse from the electronic health record in trauma patients: development and internal validation. Journal of the American Medical Informatics Association 2019;26(3):254-261. [doi: 10.1093/jamia/ocy166 PMCID: PMC6657384] [Medline: 30602031]
76. Coppersmith G, Hilland C, Frieder O, Leary R. Scalable mental health analysis in the clinical whitespace via natural language processing. 2017 Presented at: IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); 2017; Orlando p. 393-396. [doi: 10.1109/BHI.2017.7897288]
77. Coppersmith G, Leary R, Crutchley P, Fine A. Natural language processing of social media as screening for suicide risk. Biomed Inform Insights 2018;10:-11. [doi: 10.1177/1178222618792860] [Medline: 30158822]
78. Zhong QY, Mittal LP. Use of natural language processing in electronic medical records to identify pregnant women with suicidal behavior: towards a solution to the complex classification problem. European Journal of Epidemiology 2019 Feb;34(2):153-162. [doi: 10.1007/s10654-018-0470-0] [Medline: 30535584]
79. Yazdavar AH. Semi-supervised approach to monitoring clinical depressive symptoms in social media. 2017 Jul Presented at: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pages 1191-1198; 2017; Los Alamitos CA. [doi: 10.1145/3110025.3123028]
80. Dai HJ, Su ECY, Uddin M, Jonnagaddala J. Exploring associations of clinical and social parameters with violent behaviors among psychiatric patients. Journal of Biomedical Informatics 2017;75:149-159. [doi: 10.1016/j.jbi.2017.08.009]
81. Hart KL. What do patients learn about psychotropic medications on the web? A natural language processing study. Journal of affective disorders 2020 Jan 1;260:366-371. [doi: 10.1016/j.jad.2019.09.043] [Medline: 31539672]
82. Glauser T, Santel D, DelBello M, Faist R, Toon T, Clark P, et al. Identifying epilepsy psychiatric comorbidities with machine learning. Acta Neurologica Scandinavica 2020 May;141(5):388-396. [doi: 10.1111/ane.13216] [Medline: 31889296]
83. Guntuku SC. Studying expressions of loneliness in individuals using twitter: an observational study. BMJ Open 2019 Nov 04;9(11):e030355. [doi: 10.1136/bmjopen-2019-030355] [Medline: 31685502]
84. Sharma B. Validation of an alcohol misuse classifier in hospitalized patients. Alcohol (Fayetteville, N.Y.) 2020 May;84:49-55. [doi: 10.1016/j.alcohol.2019.09.008] [Medline: 31574300]
85. Barroilhet SA. Characterizing DSM-5 and ICD-11 personality disorder features in psychiatric inpatients at scale using electronic health records. Psychological Medicine 2020 Oct;50(13):2221-2229. [doi: 10.1017/S0033291719002320] [Medline: 31544723]
86. Topaz M, Murga L, Bar-Bachar O, Cato K, Collins S. Alcohol and substance abuse status from clinical notes: The added value of nursing data. Studies in health technologies and informatics 2019 Aug 21;264:1056-1060. [doi: 10.3233/SHTI190386] [Medline: 31438086]
87. Ambalavan AK, Moulahi B, Azé J, Bringay S. Unveiling Online Suicide Behavior: What Can We Learn About Mental Health from Suicide Survivors of Reddit? Studies in health technologies and informatics 2019 Aug;264:50-54. [doi: 10.3233/SHTI190181] [Medline: 31437883]
88. Bittar A, Velupillai S, Roberts A, Dutta R. Text classification to inform suicide risk assessment in electronic health records. Studies in health technologies and informatics 2019 Aug;264:40-44. [doi: 10.3233/SHTI190179] [Medline: 31437881]
89. Yan H, Fitzsimmons-Craft EE. Automatic detection of eating disorder-related social media posts that could benefit from a mental health intervention. The International Journal of Eating Disorders 2019 Oct;52(10):1150-1156. [doi: 10.1002/eat.23148] [Medline: 31381168]
90. Danielsen AA. Predicting mechanical restraint of psychiatric inpatients by applying machine learning on electronic health data. Acta Psychiatrica Scandinavica 2019 Aug;140(2):147-157. [doi: 10.1111/acps.13061] [Medline: 31209866]
91. Mezuk B, Ko TM. Suicide Among Older Adults Living in or Transitioning to Residential Long-term Care, 2003 to 2015. JAMA Network Open 2019 Jun;2(6):e195627. [doi: 10.1001/jamanetworkopen.2019.5627] [Medline: 31199445]

92. Parthipan A. Predicting inadequate postoperative pain management in depressed patients: A machine learning approach. PLoS ONE 2019 Feb 06;14(2):01-13. [doi: 10.1371/journal.pone.0210575] [Medline: 30726237]
93. Vaucheret Paz E. Sentiment Analysis in Children with Neurodevelopmental Disorders in an Ingroup/Outgroup Setting. Journal of Autism and Developmental Disorders 2020 Jan;50(1):162-170. [doi: 10.1007/s10803-019-04242-3] [Medline: 31571066]
94. Van Le D, Montgomery J, Kirkby KC. Risk prediction using natural language processing of electronic mental health records in an inpatient forensic psychiatry setting. Journal of Biomedical Informatics 2018 Oct:49-58. [doi: 10.1016/j.jbi.2018.08.007] [Medline: 30118855]
95. Corcoran CM, Carrillo F, Fernandez-Slezak D, Bedi G, Klim C, Javitt DC, et al. Prediction of psychosis across protocols and risk cohorts using automated language analysis. World Psychiatry 2018 Feb;17(1):67-75. [doi: 10.1002/wps.20491] [Medline: 29352548]
96. Moessner M, Feldhege J, Wolf M, Bauer S. Analyzing big data in social media: Text and network analyses of an eating disorder forum. International Journal of Eating Disorders 2018 Jul;51(7):656-667. [doi: 10.1002/eat.22878] [Medline: 29746710]
97. Takano K, Ueno M, Mori M, Nishiguchi Y, Raes F. Unraveling the linguistic nature of specific autobiographical memories using a computerized classification algorithm. Behavior Research Methods 2017 Jun;49(3):835-852. [doi: 10.3758/s13428-016-0753-x] [Medline: 27338931]
98. Hoogendoorn M, Berger T, Schulz A, Stolz T, Szolovits P. Predicting social anxiety treatment outcome based on therapeutic email conversations. IEEE Journal of Biomedical and Health Informatics 2017 Sep 21;21(5):1449-1459. [doi: 10.1109/JBHI.2016.2601123] [Medline: 27542187]
99. Tung C, Lu W. Analyzing depression tendency of web posts using an event- driven depression tendency warning model. Artificial Intelligence in Medicine 2016 Jan:53-62. [doi: 10.1016/j.artmed.2015.10.003] [Medline: 26616420]
100. Patel R, Jayatilleke N, Broadbent M, Chang CK, Foskett N, Gorrell G, et al. Negative symptoms in schizophrenia: a study in a large clinical sample of patients using a novel automated method. BMJ Open 2015 Sep 7;5(9):7619. [doi: 10.1136/bmjopen-2015-007619] [Medline: 26346872]
101. Iqbal E, Mallah R, Jackson RG, Ball M, Ibrahim ZM, Broadbent M, et al. Identification of adverse drug events from free text electronic patient recordsinformation in a large mental health case register. PLoS ONE. e013 2015;10(8):4208. [doi: 10.1371/journal.pone.0134208]
102. Patel R, Lloyd T, Jackson R, Ball M, Shetty H, Broadbent M, et al. Mood instability is a common feature of mental health disorders and is associated with poor clinical outcomes. BMJ Open 2015 May 21;5(5):7504. [doi: 10.1136/bmjopen-2014-007504] [Medline: 25998036]
103. Sohn S, Kocher JPA, Chute CG, Savova GK. Drug side effect extraction from clinical narratives of psychiatry and psychology patients. J Am Med Inform Assoc 2011 Dec:i144-9. [doi: 10.1136/amiajnl-2011-000351] [Medline: 21946242]
104. Yu LC, Chan CL, Lin CC, Lin IC. Mining association language patterns using a distributional semantic model for negative life event classification. Journal of Biomedical Informatics 2011 Feb 01;44(4):509-518. [doi: 10.1016/j.jbi.2011.01.006] [Medline: 21292030]
105. Miller WR, Rollnick S. Motivational Interviewing, Helping People Change. New-York: Guilford Publications; 2012:-482.
106. Mowery D, Smith H, Cheney T, Stoddard G, Coppersmith G, Bryan C, et al. Understanding depressive symptoms and psychosocial stressors on twitter: A corpus-based study. J Med Internet Res 2017 Feb;19(2):e48. [doi: 10.2196/jmir.6895] [Medline: 28246066]
107. Insel T, Cuthbert B, Garvey M, Heissen R, Pine DS, Quinn K, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am J Psychiatry 2010 Jul;167(7):748-751. [doi: 10.1176/appi.ajp.2010.09091379] [Medline: 20595427]
108. Stoyanov D, Telles-Correia D, Cuthbert BN. The Research Domain Criteria (RDoC) and the historical roots of psychopathology: A viewpoint. Eur. psychiatr 2019 Jan 21;57:58-60. [doi: 10.1016/j.eurpsy.2018.11.007] [Medline: 30677549]
109. Ross J, Neylan T, Weiner M, Chao L, Samuelson K, Sim I. Towards constructing a new taxonomy for psychiatry using self-reported symptoms. Stud Health Technol Inform 2015;216:736-740. [Medline: 26262149]
110. Vilar A, Pérez-Sola V, Blasco MJ, Pérez-Gallo E, Ballerster Coma L, Batlle Vila S, et al. Translational research in psychiatry: The research domain criteria project (RDoC). Rev Psiquiatr Salud Ment 2018 Jun;12(3):187-195. [doi: 10.1016/j.rpsm.2018.04.002] [Medline: 29941228]
111. Deo RC. Machine Learning in Medicine. Circulation 2015 Nov;132(20):1920-1930. [doi: 10.1161/CIRCULATIONAHA.115.001593] [Medline: 26572668]
112. Gholami-Kordkheili F. The impact of social media on medical professionalism: a systematic qualitative review of challenges and opportunities. Journal of Medical Internet Research 2013 Aug;15(8):e184. [doi: 10.2196/jmir.2708] [Medline: 23985172]
113. Fagherazzi G. Digital diabetes: Perspectives for diabetes, prevention and management and research. Diabetes & Metabolism 2019 Sep;45(4):322-329. [doi: 10.1016/j.diabet.2018.08.012] [Medline: 30243616]
114. Sassolas M. Défense de la psychiatrie clinique. Paris: Eres; 2014.

115. Martinez-Martin N. Is It Ethical to Use Prognostic Estimates from Machine Learning to Treat Psychosis? AMA Journal of Ethics 2018 Sep;20(9):e804-811. [doi: 10.1001/amajethics.2018.804] [Medline: 30242810]

Abbreviations

AI: artificial intelligence
ASD: autism spectrum disorder
CEGS N-GRID: Centers of Excellence in Genomic Science Neuropsychiatric Genome-Scale and RDoC Individualized Domains
cTAKES: clinical Text Analysis and Knowledge Extraction System
CUI: concept unique identifier
DSM: Diagnostic and Statistical Manual of Mental Disorders
EHR: electronic health record
EMR: electronic medical record
LDA: latent Dirichlet allocation
LogiR: logistic regression
LSA: latent semantic analysis
ML: machine learning
NLP: natural language processing
NN: neural network
POS: part of speech
PTSD: posttraumatic stress disorder
RDoC: Research Domain Criteria
RF: random forest
SVM: support vector machine
UMLS: unified medical language system

Edited by G Eysenbach; submitted 31.07.19; peer-reviewed by M Sharma, H Mufti, M del Pozo Banos, V Curcin, G Hu, D Leightley, SGS Shah, N Miyoshi, F Lanfranchi; comments to author 21.10.19; revised version received 18.04.20; accepted 02.10.20; published 04.05.21

Please cite as:
Le Glaz A, Haralambous Y, Kim-Dufor DH, Lenca P, Billot R, Ryan TC, Marsh J, DeVylder J, Walter M, Berrouiguet S, Lemey C
Machine Learning and Natural Language Processing in Mental Health: Systematic Review
J Med Internet Res 2021;23(5):e15708
URL: https://www.jmir.org/2021/5/e15708
doi: 10.2196/15708
PMID:

Review

Information and Communication Technology Use in Suicide Prevention: Scoping Review

Jessica Rassy[1,2,3,4*], PhD; Cécile Bardon[1,5*], PhD; Luc Dargis[1*], MSc; Louis-Philippe Côté[1,5*], MSc; Laurent Corthésy-Blondin[1,5*], MSc; Carl-Maria Mörch[1,5,6,7*], PhD; Réal Labelle[1,2,5,8], PhD

[1]Center for Research and Intervention on Suicide, Ethical Issues and End-of-Life Practices, Université du Québec à Montréal, Montréal, QC, Canada
[2]Research Center, Institut universitaire en santé mentale de Montréal, Montréal, QC, Canada
[3]School of Nursing, Université de Sherbrooke, Longueuil, QC, Canada
[4]Quebec Network on Nursing Intervention Research, Montréal, QC, Canada
[5]Department of Psychology, Université du Québec à Montréal, Montréal, QC, Canada
[6]Algora Lab, Université de Montréal, Montréal, QC, Canada
[7]Mila, Quebec Artificial Intelligence Institute, Montréal, QC, Canada
[8]Department of Psychiatry, Université de Montréal, Montréal, QC, Canada
[*]these authors contributed equally

Corresponding Author:
Réal Labelle, PhD
Department of Psychology
Université du Québec à Montréal
CP 8888, succ Centre-Ville
Montréal, QC, H3C 3P8
Canada
Phone: 1 514 987 4184
Email: labelle.real@uqam.ca

Abstract

Background: The use of information and communication technology (ICT) in suicide prevention has progressed rapidly over the past decade. ICT plays a major role in suicide prevention, but research on best and promising practices has been slow.

Objective: This paper aims to explore the existing literature on ICT use in suicide prevention to answer the following question: what are the best and most promising ICT practices for suicide prevention?

Methods: A scoping search was conducted using the following databases: PubMed, PsycINFO, Sociological Abstracts, and IEEE Xplore. These databases were searched for articles published between January 1, 2013, and December 31, 2018. The five stages of the scoping review process were as follows: identifying research questions; targeting relevant studies; selecting studies; charting data; and collating, summarizing, and reporting the results. The World Health Organization suicide prevention model was used according to the continuum of universal, selective, and indicated prevention.

Results: Of the 3848 studies identified, 115 (2.99%) were selected. Of these, 10 regarded the use of ICT in universal suicide prevention, 53 referred to the use of ICT in selective suicide prevention, and 52 dealt with the use of ICT in indicated suicide prevention.

Conclusions: The use of ICT plays a major role in suicide prevention, and many promising programs were identified through this scoping review. However, large-scale evaluation studies are needed to further examine the effectiveness of these programs and strategies. In addition, safety and ethics protocols for ICT-based interventions are recommended.

(J Med Internet Res 2021;23(5):e25288) doi: 10.2196/25288

KEYWORDS

suicide prevention; information and communication technology; scoping review; mobile phone

Introduction

Background

Information and communication technology (ICT) has been used for suicide prevention over the past decade. Moreover, there is a growing body of evidence supporting the use of ICT in the development of promising suicide prevention practices [1,2]. ICT can be used to screen individuals at risk of suicide on the web; offer information and help regarding suicidal thoughts and behavior; and offer web-based assessment, interventions, and follow-up [1-6]. The use of ICT widens accessibility to hard-to-reach individuals who do not always seek help in person and offers treatment opportunities to communities with lower access to care, such as rural communities. The use of ICT in suicide prevention can also help professionals offer better care to their patients by combining multiple approaches, such as using mobile apps to monitor symptoms or providing web-based therapeutic programs [7,8].

Many opportunities arise when using ICT to expand suicide prevention strategies. However, the pace at which ICT is advancing makes it difficult to keep up with, especially from a research perspective. To make better use of these different web-based suicide prevention strategies, a better understanding of these different uses of ICT is necessary. We identified 3 major questions. First, a better understanding of how to emphasize the technical aspects of ICT regarding the development, maintenance, privacy, and life cycle of the technology would help make better choices regarding which ICT to use in what context. For example, machine learning offers many possibilities for identifying at-risk individuals, but major technical and ethical considerations must be described and taken into account. Second, we need to improve our understanding of ICT use in suicide prevention. To support adequate decision making, it is important to know whether ICT-based suicide prevention strategies reach different clienteles or if we are reaching the same individuals differently. Third, a better understanding of the structure and efficacy of various types of current web-based assessments and interventions is also important. For instance, can artificial intelligence (AI) and machine learning be used to optimize and accelerate the assessment of individuals at risk? Is a suicide crisis intervention by chat as effective as talking over the phone or in person? Are web-based cognitive behavioral therapy (CBT) programs as effective for suicide prevention as face-to-face CBT programs? Before shifting toward using and recommending the use of these different ICT-based assessments and interventions, it should be noted that their effectiveness has yet to be demonstrated in the literature.

Current Study

The available literature often refers to a specific type of ICT as opposed to a general overview of existing ICT-based strategies in mental health and suicide prevention. More precisely, existing literature reviews have focused on specific ICTs such as smartphone tools [1], web-based interventions [3], mobile apps [4], and social media [5] or specific age groups such as youth [6]. Based on this background and a rapidly evolving corpus of research, we set out to review the evidence supporting the use of all types of ICTs in different levels of suicide prevention strategies. More precisely, we carried out a scoping review to identify the best and most promising practices for ICT use at all levels of suicide prevention, which are described by the World Health Organization as the universal (entire population), selective (specific subpopulations), and indicated (high-risk individuals) levels [9]. ICT was defined as all materials, software, or services used to collect, process, and transmit information; this includes electronic, computer, telecommunications, multimedia, and internet technologies [10]. AI was defined as all theories and techniques used to develop machines capable of simulating intelligence [7].

Methods

Scoping Review Framework

We used the 5-stage scoping framework developed by Arksey and O'Malley [11] and adapted by Levac et al [12] to review peer-reviewed publications. These stages are (1) identifying research questions; (2) identifying relevant studies; (3) selecting studies; (4) charting data; and (5) collating, summarizing, and reporting results. A scoping review allows researchers to explore the available data on various aspects of a theme instead of just one [11,12]. This provides an overview of the relevant literature that helps to identify the various ICT and suicide prevention strategies examined by researchers since 2013 and to document their effects. The Center for Research and Intervention on Suicide, Ethical Issues, and End-of-Life Practices decided to better understand the use of ICT in suicide prevention as of 2013 following the publication of the book *Suicide prevention and new technologies: Evidence-based practice* [13]. This book provides an overview of new technologies in suicide prevention and demonstrates the need for further research in this area. As ICT tools, contents, and use evolve rapidly, a 5-year timeframe seems relevant to address the current state of ICT preventive practices and research. For example, major social networks have changed their policies toward content related to suicidality. During the 2019 International Suicide Prevention Day, Facebook revised its policy of preventing self-harm and suicide by banning graphic representations of suicidality, which could change the level of exposure of individuals vulnerable to suicide. At the same time, the rapid development of scientific knowledge and technology has been noted many times. For example, a review of the scientific literature published in 2013 identified 5 scientifically evaluated mobile apps for child and adolescent mental health [14], whereas Grist et al [15] identified 19 more. In addition, web-based platforms such as Parler or 4Chan were not as widely known and used 10 years ago. Including older technologies and neglecting to consider the rapid content change of internet content over the years may skew results toward obsolete prevention strategies.

Summary of Search Strategy

Identifying the Research Question

The research question identified for this scoping review aims to guide future program development and implementation by establishing an inventory of strategies using ICT that have been the subject of research in recent years. Following the PRISMA

(Preferred Reporting Items for Systematic Reviews and Meta-Analyses) recommendations, the research question was developed using the PICO (population, intervention, comparison, and outcomes) conceptual tool [16]. Note that an adaptation of PICO was made through an amalgamation with the SPIDER (sample, phenomenon of interest, design, evaluation, research type) tool by Methley et al [17] to facilitate the consideration of the psychosocial nature of the research object. Consequently, the research question was "How information and communication technology (ICT) can be used in intervention and prevention of suicide among the general population, people at risk, or suicidal people of all ages and conditions combined to contribute to the reduction of any indicators relating to suicidality?" This question was selected to promote the instrumental use of the research results [18].

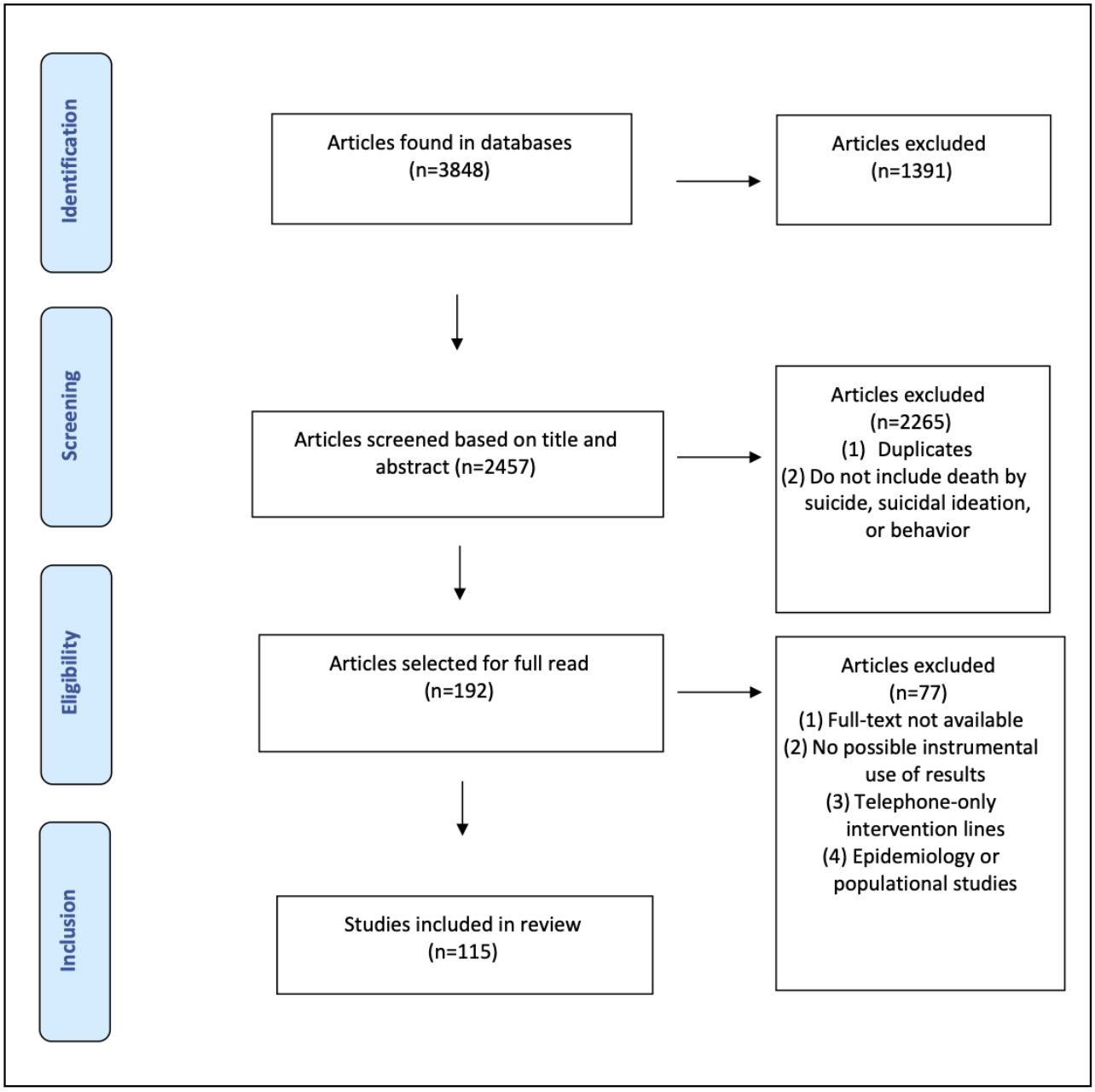
Identifying Relevant Studies

The search strategy consisted of identifying key concepts related to ICT-based suicide prevention and intervention, developing a provisional syntax specific to each database for pretesting, and determining the definitive syntax based on the pretest results. The keywords used to construct research syntaxes were extracted from natural (everyday vocabulary) and controlled (indexer terms) language terms and based on the central concepts associated with the research questions. The researched concepts were suicidal behavior, ICT (AI, machine learning, application, social network, and mobile technology), intervention (information, education, prevention, decision making, and intervention), and program evaluation. The final syntax based on these keywords and used with each database searched is presented in Multimedia Appendix 1. PubMed, PsycINFO, Sociological Abstracts, and IEEE Xplore were searched for publications from January 1, 2013, to December 31, 2018. The selection of each database was carefully thought out by the research team (2 senior researchers in suicidology, 2 ICT experts, 1 postdoctoral student, 1 doctoral student, and 1 specialized librarian who provided support) to include data in the medical field, psychology, social and behavioral sciences, and ICT. The number of databases chosen was in line with that of other reviews that focused on specific ICTs and mental health [5,19]. In addition, the choice of databases was based on the following 2 criteria: the number of documents they contain and the degree of overlap of the content indexed in each database [20]. The final choice was made to select high quantities of indexed documents in databases and databases with less content overlap to maximize the completeness of the retrieval process.

Selecting Studies

The final search syntax identified 3848 publications that were imported for EndNote selection. From these, 1391 duplicates were removed. The remaining 2457 publications were then sorted according to the inclusion and exclusion criteria applied by 5 research assistants against titles and abstracts only. Interrater reliability was assessed for 100 publications. Agreement reached 89% and a Cohen κ statistic of 0.704 was achieved, which is considered substantial by Cohen [21]. The studies covered in our review used various methods (qualitative, quantitative, and mixed methods) and were drafted in English or French. They had to provide original empirical or descriptive data. Literature reviews, editorials, theoretical articles, and studies focused only on ethical or legal issues surrounding ICT-based suicide prevention were excluded. The included articles had to directly address suicide prevention strategies using ICT. All studies that used death by suicide, suicidal ideation, or suicide behavior as outcomes were considered. The PRISMA flow chart [16] of our study presented in Figure 1 provides an overview of our study selection process.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.



Charting, Collating, Summarizing, and Reporting Data

A form was developed in Microsoft Excel to retrieve pertinent information from selected qualitative, quantitative, and mixed methods studies to optimize data charting. A research team tested the form before use. It included 10 items: title, objectives, research design, instruments of measure, type of suicide prevention (universal, selective, or indicated), technology category, participant characteristics, results associated with effects and benefits regarding lowering suicidal behavior, clinical and scientific contributions, and recommendations. The research team thoroughly examined 5 independent studies with 10 items and then compared the results. Following this analysis, precision was applied to the exclusion criteria of the studies. Verbal telephone interventions were excluded from this review. Considering the large quantity of data available on suicide prevention telephone interventions [22-24], studies addressing these types of interventions were excluded from this scoping review. However, interventions based on text messaging and smartphone apps were included as new ICT-based interventions relevant to this review. After criteria validation, a summary of pertinent information was prepared for each study included. No systematic methodological quality evaluation was carried out in accordance with the scoping review methodology [11].

Results

General Information

Our search of PubMed, PsycINFO, Sociological Abstracts, and IEEE Xplore yielded 3848 articles. After removing 1391 duplicates and 2265 articles based on a perusal of title and abstract, 192 texts were read in full for the final screening. This allowed us to remove an additional 77 articles, leaving 115 articles for the scoping review. We also noticed through our

analysis that the number of publications increased in 2016 and 2017, and the scientific articles focused more on selective and indicated suicide prevention strategies.

Classification of Interventions' Efficacy

In the context of this scoping review, we aim to provide a general portrait of existing suicide prevention strategies based on ICT. It does not aim to determine the quantitative effectiveness of such interventions. Therefore, a prevention strategy is considered effective when it is based on rigorous theory and when it has been evaluated by a minimum of 2 studies with a quasi-experimental approach [25]. Typically, in traditional systematic reviews, identifying *evidence-based practices* requires a weighting of the methodological quality of primary studies, unlike the scoping review [12]. If the relevant studies are randomized trials, the body of evidence begins with high certainty. If the relevant studies are observational, the body of evidence begins with low certainty [26]. Certain strategies may, however, be presented as *promising* when at least one observational or randomized study supports its efficacy. Therefore, we considered all interventions that have demonstrated effectiveness in any type of research design (qualitative, quantitative, or mixed studies).

Universal Suicide Prevention Strategies

In total, 10 of the studies selected referred to universal suicide prevention strategies (Table 1). These addressed 2 main program categories: (1) health promotion and suicide prevention through the use of educational websites and (2) health promotion and suicide prevention through awareness campaigns and social media psychoeducation. These programs were accessible to all participants. The types of websites identified were information-based, interactive, forums, and chats. Social media platforms included Facebook, Twitter, and personal and professional blogs. In general, the effects of these programs on perceptions and knowledge have been poorly evaluated. Universal suicide prevention strategies included the *SUPREME* (Suicide Prevention through Internet and Media-Based Mental Health Promotion) project [27], the *Storytelling* project [28], the *It Gets Better* project [29], the *Live Through This* project [30], and *Media-Based Prevention Messages* [31]. Although these different universal strategies could have some effects on the negative emotions of participants [32], they primarily play a potential role in reducing the stigma associated with suicide [30,33] and suicidal ideation and behavior [27,30,32] and increasing web-based suicide prevention knowledge [31,34,35], mental health literacy [32,33], and web-based help seeking [33,34]. Table 1 presents a description of the studies that referred to universal suicide prevention strategies.

Table 1. Universal suicide prevention strategies.

Type of program	Identified studies	Objectives	Programs, n	ICT[a] used	Targeted population	Examples	Results and comments
Health promotion and suicide prevention through the use of educational websites	[27,28,30,32]	• Improve mental health and well-being • Educate and increase awareness of young people of mental health and suicide prevention • Prevent stigma • Encourage help seeking and service use by young people • Promote protective factors	2	• Websites • Interactive modules • Forums	• Young people • Young adults • At-risk groups	"SUPREME[b]" project [27] and "Live Through This" [30]	Results in terms of reducing suicidal ideation and behavior, increasing mental health literacy, and increasing web-based help seeking seemed promising.
Health promotion and suicide prevention through awareness campaigns and social media psychoeducation	[29,31,33-36]	• Increase social media users' awareness of the existence of and issues associated with suicide and its prevention	4	• Presenting, broadcasting and sharing messages on social media	• Young people • At-risk groups	"It Gets Better" project [29] and "Media-Based Prevention Messages" [31]	Results in terms of reducing suicidal ideation and behavior and improving knowledge, attitudes, and asking for help on the web seemed promising. However, further research was needed.

[a]ICT: information and communication technology.

[b]SUPREME: Suicide Prevention through Internet and Media-Based Mental Health Promotion.

Selective Suicide Prevention Strategies

We identified 53 studies that dealt with selective suicide prevention strategies (Table 2). From these, 9 different selective suicide prevention program categories emerged, and each had multiple different outcomes. These strategies ranged from identifying people at risk to web-based self-management and training programs. Some were designed to be used alone, whereas others were part of a larger intervention program involving direct contact between suicidal individuals and suicide prevention resources. The programs targeted distressed individuals and groups at risk, that is, health professionals. ICT use varied according to the desired outcome. For example, algorithms were used to identify individuals at risk for suicide on the web, as advertised by Google AdWords, which is now known as Google Ads [37,38]; interactive websites were used to support the assessment and management of mental health disorders in collaboration with a mental health professional [39]; and training modules, web-based exercises, and multimedia presentations were used to enhance knowledge, understanding, and attitudes regarding suicide prevention and intervention [40-42]. Programs using algorithms to identify at-risk individuals seemed most promising for identifying people at risk for suicide versus those not at risk [37,38,43-58]. These algorithms analyzed suicide risk on the basis of different elements, including speech and linguistic characteristics, medical notes, search engine ad clicks, and profiles from web-based chat sessions or social media [37,38,43-58]. Although these algorithms helped identify individuals at risk for suicide, a clinical application of these algorithms is yet to be developed, as little is known about the effectiveness of these programs in increasing actual web-based and in-person help-seeking behaviors.

Table 2. Selective suicide prevention strategies.

Type of program	Identified studies	Objectives	Programs, n	ICT[a] used	Targeted population	Examples	Results and comments
Identifying at-risk individuals through automatic analysis of their linguistic characteristics during face-to-face consultations	[44-47]	• Identify individuals at risk for suicide and qualify risk level based on speech characteristics	2	• Analysis of characteristics from observation of spoken language	• Individuals presenting risk factors at assessment interviews	Analysis of speech and sound characteristics program [45-47]	This technology seems effective in distinguishing suicidal from nonsuicidal individuals. A clinical use of this technology is yet to be developed.
Identifying at-risk individuals through automatic analysis of medical file notes and/or research data	[54-57]	• Identify individuals at risk for suicide and qualify risk level based on data (medical file or research data) from clinical observations by clinicians or researchers	1	• Analysis of written language	• Individuals presenting risk factors during assessment interviews	The "SafetyNet" program [56]	This technology seems effective in distinguishing suicidal from nonsuicidal individuals. A clinical use of this technology is yet to be developed.
Identifying at-risk individuals through automatic analysis of their linguistic characteristics during chat sessions or on written forms	[43,58]	• Identify individuals at high risk for suicide based on characteristics of their words or responses	0	• Analysis of written language	• Individuals presenting risk factors during written exchanges	Suicide meaning making applied in specific contexts [58]	This technology seems promising. A clinical use of this technology is yet to be developed.
Identifying at-risk individuals through automatic analysis of their linguistic characteristics on social media	[48-53,59-62]	• Examine writings of social media users to automatically identify individuals at risk for suicide and offer them proactive help	4	• Analysis of written language	• Social media users	Program to screen suicidal individuals based on tweets [52,59]	This technology is innovative. Further research is needed.
Identifying at-risk individuals through targeted ads on search engines	[37,38]	• Identify suicidal individuals through their queries on search engines • Suggest help resources	2	• Search engine algorithm and targeted ads	• General public	Web-based sentinel program [37]	The number of clicks on targeted ads serves as an indicator. There are too few results to date to evaluate these programs' effectiveness in increasing the web-based help-seeking behaviors of suicidal individuals.
Web-based identification programs tailored to different groups	[63-69]	• Identify vulnerable groups using mailing lists or websites to direct them toward help resources	3	• Web-based standardized risk assessment questionnaires with response algorithms adapted to assessment results	• Individuals at risk for suicide	EMPATHY[b] [63] and HEAR[c] for nurses [69]	This type of program allows identifying individuals at risk who have no contact with health services. Utilization rates for proposed resources seem encouraging.

Type of program	Identified studies	Objectives	Programs, n	ICT[a] used	Targeted population	Examples	Results and comments
Web-based mental health self-evaluation programs	[39,70]	• Support evaluation and management of mental health problems in conjunction with a mental health professional	1	• Interactive website • Web-based evaluation tool • Algorithm tool for redirecting to resources • Development of a web-based personalized intervention plan	• Individuals receiving support from a health professional	myGRaCE decision support system [39] and YouthCHAT [70]	Improves patient engagement in understanding their situation and in managing their treatment.
Web-based mental health self-management program with measures of impact on suicide risk	[71-81]	• Facilitate web-based mood self-management and improve quality of life through a cognitive behavioral intervention	5	• Interactive website • Psychoeducation module • Multimedia presentation • Web-based exercises • Discussion and consultation forum	• Individuals with low-intensity mental health problems	MindSpot Clinic [77]	Program membership rates are relatively low. Individuals tend more often to experience a decrease in suicidal ideation and depression symptoms. Programs are offered alone or in conjunction with clinical follow-up by a professional. Results are hard to compare.
Web-based suicide prevention training programs	[40-42,82-87]	• Improve knowledge, attitudes and suicide prevention practices of professionals and sentinels	5	• Website • Training modules • Multimedia demonstration • Web-based exercises	• Professionals • Sentinels	Question, Persuade, Refer, and Treat program [40,84]	These programs allow improving knowledge and attitudes. They seem less effective in changing intervention practices if not offered along with certain forms of face-to-face practices.

[a]ICT: information and communication technology.

[b]EMPATHY: Empowering a Multimodal Pathway Toward Healthy Youth.

[c]HEAR: Healer Education Assessment and Referral.

Using ICT to identify individuals at risk for suicide in school (from elementary school to medical school) or community settings has been associated with increased mental health literacy, help-seeking behavior, and use of help resources [63-65]. For example, in the Health Education Assessment and Referral (HEAR) program, medical students anonymously responded to a web-based questionnaire on suicide risk and various mental health issues. The program allowed identifying students at risk for suicide and referring them to a medical school psychiatrist or psychologist through a web-based platform. As a result, the use of medical schools' mental health services increased from 11.5% to 15% over 4 years [64].

Our review also identified web-based self-assessment and self-management programs. For example, myGRaCE is a decision-making support system that combines service user self-assessment and practitioner expertise by comparing the user's self-assessment against a practitioner's assessment [39]. Most of the participants in this study agreed that myGRaCE helped them assess their personal security, understand what puts them in danger, and what areas they should change in their lives. Self-assessment programs are also a way to engage young people and adults by helping them understand their situation and identify ways to take control of their health [39,70]. As for web-based self-management programs, results showed that many programs could, in some cases, significantly reduce suicidal ideation [71-74] and increase chances of resorting to mental health interventions [75,76]. However, the rate of adherence to these types of programs remains low. Examples of these include MoodGYM [71], MindSpot Clinic [77], different types of internet-based cognitive behavioral therapy (iCBT), Thrive [73], the Sadness program [78], and CATCH-IT (Competent Adulthood Transition with Cognitive, Behavioral,

Humanistic and Interpersonal Training) [74]. These programs were offered either alone or in combination with clinical follow-up. Users seemed to appreciate the mobile apps used for self-management and suicide prevention training. Many web-based suicide prevention programs have been tested for their efficacy. These programs were intended for gatekeepers who work with adolescents [82], gatekeepers in school settings [83], mental health professionals [40], health professionals in general [41,42,84], Veterans Affairs providers [85], and graduate students [86]. In some cases, face-to-face training developed more knowledge and suicide prevention skills than web-based training [84]. In addition, an increase in suicide prevention knowledge was often observed within the first months post training but often decreased over time [82]. Details of the studies that referred to selective suicide prevention strategies are presented in Table 2.

Indicated Suicide Prevention Strategies

Regarding the indicated suicide prevention strategies, we selected 52 studies for our review and identified 9 program categories. These programs, presented in Table 3, were offered by health professionals or psychosocial help services and covered suicide risk assessment triage and monitoring, crisis intervention, low-intensity psychological interventions, psychotherapy for individuals at risk for suicide, and technological tools to support face-to-face interventions. The ICTs used for these programs included classification algorithms, chats and text messages, mobile apps, and interactive clinical intervention websites. Examples of these programs include the Mental Health eClinic [55], a web-based assessment and self-management program; Reframe-IT [88], an internet-based CBT program for high school students at risk for suicide; MYPLAN [89], a mobile phone safety plan app for supporting people at risk for suicide; dialectical behavioral therapy Coach Mobile [90], a CBT support intervention mobile phone app; and RAFT (Reconnecting After a Suicide Attempt) [91], a brief web-based outreach intervention for people who attempted suicide and who lost contact with services. In general, ICT-based self-assessment tools have been reported to be as effective in person as in hardcopy format in identifying suicide risk warning signs [92].

Table 3. Indicated suicide prevention strategies.

Type of program	Identified studies	Objectives	Programs, n	ICT[a] used	Targeted population	Examples	Results and comments
Development of web-based suicide risk assessment tools	[92-97]	• Propose suicide risk assessment tools administered electronically in clinical contexts	4	• Electronic clinical instrument • Response selection algorithm based on assessment results	• Individuals at risk for suicide in contact with health services	Web-based Columbia Suicide Severity Rating Scale [93]	Electronic assessment tools seem as effective as hardcopy versions. They allow internet users at times to reveal their suicidal behaviors more easily
Use of artificial intelligence and machine learning to optimize completion time for suicide risk assessment tools	[98]	• Reduce completion time for suicide risk assessment tools	1	• CAT[b]	• Individuals at risk for suicide in contact with health services	CAT [98]	This type of algorithm allows reducing the number of items needed to assess suicide risk.
Use of triage systems by ICT to assess suicide risk	[99]	• Improve triage of individuals at risk for suicide who use health services	1	• On-lining of clinical suicide risk assessment tools • Risk-level classification algorithm based on scores • Targeted offer of resources • Automatic alert to the clinical team	• Individuals at risk for suicide in contact with health services	Mental health eClinic [99]	This type of program allows identifying young people at risk for suicide and offering them clinical treatment more rapidly.
Crisis intervention via text messaging or internet chatting	[68,69,84,85]	• Intervene in a crisis situation via text messaging or internet chatting	3	• Text messaging • Internet chatting • Computerized system for processing text messages and chat exchanges	• Young people are the priority target group of these interventions. They can, however, be used to reach other age groups	Kids Help Phone LIVECHAT Project [100] and RAFT[c] [91]	Text messages and internet chat interventions seem a viable alternative to telephone interventions for certain groups. However, intervention strategies adapted to these modes of communication need to be developed.
Use of web-based publications by patients as intervention support material	[101,102]	• Assess suicide risk and intervene using a person's web-based discourse	1	• Content of posts written on social networks	• Suicidal individuals receiving mental health services	Patient's social networking sites as a clinical tool [101]	The use of content written on social networks by patients allows a better understanding of the situation and helps perform a suicide risk assessment, especially in the case of patients who deny such behaviors.

Type of program	Identified studies	Objectives	Programs, n	ICT[a] used	Targeted population	Examples	Results and comments
Web-based management and low-intensity intervention for individuals at risk for suicide	[1,31,63,68,88,102-114]	• Intervene with individuals at risk for suicide using a cognitive behavioral approach	7	• Interactive website • Web-based exercise module • Discussion forum • Multimedia presentation	• General population • Individuals referred to programs by clinicians	Reframe-IT [88], Fitmindkit [102], EMPATHY[d] [63], and SMART[e] Mental Health Project [110,113]	These programs bring about a small decrease in suicide risk among participants.
Mobile apps used as part of treatment follow-up with suicidal individuals	[90,115-121]	• Support face-to-face clinical intervention • Support aboriginal young people by complementing face-to-face intervention	2	• Interactive mobile app • Algorithm for providing responses and offering resources based on information provided by the user	• Individuals receiving mental health services • Aboriginal young people	DBT[f] Coach Mobile [90] and AIMhi Stay Strong iPad [101]	These apps seem to help reduce the danger of suicide and the urgency of self-harm. Perceptions are positive, and apps allow reaching young people in isolated communities.
Use of web-based monitoring tools to improve psychological follow-up of individuals at risk for suicide	[107,122,123]	• Conduct regular evaluations of individuals and their symptoms to inform and adjust intervention	3	• Digitalization of screening tools • Mobile app • Contact with intervention team via automated alerts • Automated sending of text messages	• Individuals diagnosed with a mood disorder and receiving mental health services	Depression Project [123]	The use of an app facilitates disclosure of suicidal and self-harm behaviors. Regular monitoring of symptoms helps adjust intervention strategies.
Follow-up program by automatic text messages for individuals at risk for suicide	[91,124-126]	• Offer tailored follow-up following a suicide attempt to increase treatment use and reduce suicidal behaviors and self-harm	3	• Automated sending of predrafted text messages, including encouragement and appointment date reminders	• Suicidal individuals receiving mental health services	Postattempt follow-up program by text messaging [124]	Users appreciate the text messages and deem them a good way of keeping in touch with care services. Help-seeking behaviors increase, and self-harm behaviors decrease.

[a]ICT: information and communication technology.

[b]CAT: computerized adaptive testing.

[c]RAFT: Reconnecting After a Suicide Attempt.

[d]EMPATHY: Empowering a Multimodal Pathway Toward Healthy Youth.

[e]SMART: Systematic Medical Appraisal, Referral and Treatment.

[f]DBT: dialectical behavioral therapy.

In some cases, web-based self-assessment tools, as opposed to face-to-face tools, seemed to facilitate the disclosure of suicidal behavior [93]. AI has also been shown to be effective in optimizing the web-based assessment of suicide behavior by using an item response-based computer-adaptive simulation to reduce the length of a suicide risk assessment tool [98]. As for ICT-based triage systems, the use of the Synergy Online System allowed young people to complete a web-based clinical

assessment before a face-to-face or web-based clinical appointment, who were to be prioritized and contacted immediately in case of high suicide risk [99]. Web-based crisis interventions by chat or SMS text messaging have been on the rise and seem to be an alternative to phone interventions for some groups. Studies comparing chat services with SMS text messaging have shown similar results [127]. However, users of chat services, such as 113Online [128], were more likely to be at a high level of suicide crisis, had more mental health problems, were younger, and were more often to be women compared with crisis hotline users. In addition, interventions delivered through chat services were longer and more complicated, and fewer changes were observed in the individual's emotional state [127,128]. Interventions have yet to be developed and adapted to this type of technology; the fact that interventions designed to be delivered by telephone were delivered by chat or SMS text messaging without being adapted to these other media was considered a major limitation [128]. Some studies also analyzed content posts by suicidal patients on social networking sites to better assess suicide risk, especially in patients denying suicidal behavior [101,129]. Although social media content helped with better understanding a person's situation, there were many ethical concerns regarding this data collection method [129].

Self-management and iCBT programs seemed to reduce suicidal ideation. Unlike iCBT and self-management interventions addressing general mental health in the selective suicide prevention section, these interventions specifically addressed user suicide behaviors. Examples of these programs included Empowering a Multimodal Pathway Toward Healthy Youth [63], Reframe-IT [88], Fitmindkit [102], Safe conversation [103], Latitudes [103] and PrevenDep [104]. Intervention goals included developing basic problem-solving skills [88] and cognitive restructuring [105] to reduce suicide risk. Other ICT suicide prevention interventions include using computer alerts or mobile apps to apply a safety plan. These ranged from a system alerting the clinician not to forget to use the safety plan with a patient [130] to a mobile app of a personalized safety plan that patients could have on their phones at all times [89,115]. Research has demonstrated a good level of acceptability [131] for these interventions, but further evaluation of their effects on suicidal behaviors is necessary. Mobile apps have also been used to support therapeutic follow-up for suicidal patients [90,115,116], including for specific groups such as Aboriginals and Torres Strait Island Australians [117]. These mobile apps seemed to help reduce suicidal danger and self-harm [90,118], reach young people in isolated communities [132], and increase adherence to face-to-face follow-ups [119].

Discussion

Limitations

The rapid expansion of ICT use in suicide prevention has preceded the development of theoretical models to orient its methods and content and the accumulation of sufficient empirical research to indicate what is most helpful and what is not helpful. In this context, common sense has been the guiding principle, resulting in a plethora of suicide prevention activities being implemented with limited research evaluations of their effectiveness slowly following. Many of the most widely used ICT suicide prevention services have never been evaluated. Therefore, the published studies concern a small nonrandom sample of select interventions that a few researchers have been interested in studying. When little or no research has been published on ICT programs, this does not mean that they are not effective. Similarly, when there are promising empirical data on the benefits of a program, this does not mean that it is better or more useful than programs that have not been studied. Furthermore, we do not know if the *promising* results reported will stand the test of time, as more research is conducted using more rigorous research methodologies. Therefore, any conclusions drawn from the limited scientific publications must be considered preliminary and hopefully will be subject to verification in the future.

Contributions

This scoping review on ICT use in suicide prevention shows that a large number of studies have been published in the past few years [1,3,5]. Our findings shed light on the use of ICT in different types of universal, selective, and indicated suicide prevention strategies.

Amid publications with the potential to reach a large population, there are only a few publications on the use of ICT in universal suicide prevention [27-35,133]. Our findings reveal that there are around 85% fewer studies on universal prevention strategies than those on selective and indicated strategies. The 10 publications we identified describe health promotion and suicide prevention through educational websites, web-based awareness campaigns, and social media [27-35,133]. The limited empirical findings suggest that these programs may play a role in increasing general mental health literacy and the incidence of help-seeking behavior, which are associated with reduced suicidal ideation and behavior. However, the effects of these programs on perceptions and knowledge have rarely been investigated. It is important to identify which characteristics of educational websites, awareness campaigns, and postings on social media are associated with positive changes in help seeking and reductions in suicidal ideations and behaviors. Furthermore, it is essential to develop effective means for identifying sites and postings that are helpful and notifying users about or orienting them toward internet content that may be of help to them.

Selective suicide prevention strategies using ICT consist of programs that identify specific subpopulations to offer specialized support to reduce suicide risk. They particularly targeted young people and various at-risk groups (eg, the lesbian, gay, bisexual, transgender, queer, and other community; sexual minorities; and aboriginal communities) through their profiles and linguistic characteristics (eg, chat sessions, web-based written forms, and social media) or by analyzing their medical records and research data. They also include targeted web-based advertising, web-based self-assessment and self-management, and web-based training. A larger published body of research indicates that these approaches are promising, particularly in controlled environments such as schools. School programs using ICT, such as HEAR, have been shown to have a positive impact

on mental health help seeking and service use in youth [64]. Programs targeting specific subgroups of the general population have been shown to increase mental health literacy and the chances of using mental health services, but this effect seems to decrease over time [71,73,74,77,92]. Therefore, it is important to focus on the sustainability of the effects of programs that have an initial positive impact. The novelty of an intervention may be associated with greater effects. If novelty is a key feature of programs' success, then either the programs need to be constantly changed and renewed to sustain their impact or the programs need to be continually replaced by new and different activities to ensure that people will continue to be helped over time. Therefore, both approaches may be proven to be unsustainable over time. This rapid effect can also give a false sense that resorting to web-based strategies is sufficient. However, web-based help does not replace face-to-face or direct support and care from trained professionals. These different selective suicide prevention strategies should only be used in addition to face-to-face type of help and interventions. Moreover, our empirical findings show that more research is required to better distinguish between false positives and false negatives in these web-based identification techniques. It is of utmost importance to help identify at-risk groups and avoid discarding individuals who are assessed as false negatives too quickly. A false impression of security can put these individuals at a greater risk.

Where indicated suicide prevention strategies are concerned, programs vary, including using ICT in suicide risk assessment triage, monitoring suicide risk, crisis intervention activities, and psychotherapy. The programs were offered on the web only (eg, websites, mobile phone apps, and chats), but they were sometimes supported by face-to-face interventions. In some instances, they were found to be efficient (eg, shortened web-based assessments), and sometimes they were time-consuming (eg, interventions were longer in chat sessions). ICT-based and self-management interventions seem promising, as they address suicidal behaviors directly and offer alternatives for coping with suicidal thoughts by developing problem-solving skills [88,90-92]. In addition, specific at-risk groups in isolated communities where other services are not available may benefit from mobile apps used for patient follow-up.

Ethical Considerations to Address in the Future

Many of the studies included in this review raised security and ethical concerns regarding web-based suicide prevention practices. Ethical concerns range from a lack of training and web-based moderators' skills to the lack of an evidence-based framework providing guidelines for the secure use of ICT [35]. As mentioned by Robinson et al [35], it is ethically necessary to provide security protocols and a clear code of ethics for safe web-based intervention. Another ethical concern is that all individuals have the right to privacy, including web privacy [101,129]. Thus, there is reason to question whether the content analysis of web-based social media posts, emails, and other web-based sources of information is an ethical research endeavor if informed consent is not obtained from the individuals who have posted the information [101,129]. Beyond security and privacy, there are many other ethical concerns regarding web-based surveillance, informed consent, communication, controls, and disclosure [134,135]. For example, a proper ethical assessment of risks and benefits to the use of different ICT strategies is rarely, if almost never, considered in the development process. There is also a major ethical concern regarding reducing or eliminating in-person services and replacing them with insufficient web-based solutions that may appear to present better cost-effectiveness without proper in-depth assessment. With this concern, web-based interventions should always be combined with in-person formal help and intervention services. This raises many concerns and provides plenty of grounds for further research aimed at ensuring safer and more ethical use of ICT in suicide prevention.

Conclusions

As the number of studies on ICT use in suicide prevention is growing, the published literature needs to be reviewed regularly. This scoping review shows that ICT use in suicide prevention provides an interactive, personalized, readily available, and accessible approach to reach various populations for identification of at-risk individuals and to provide support. ICT may provide a sense of *being connected* to people who are otherwise isolated and reluctant to use offline services. Promising published findings on web-based intervention content includes psychoeducation and skills training. As digital help proliferates, one should consider whether this means that help-seeking and suicide prevention activities will replace traditional offline services. However, in some areas where radical changes were expected, existing modalities continue to be used (eg, individuals did not stop using in-person services when telephone phone crisis lines appeared, and these telephone lines reached a different audience and complemented existing services). The extent to which ICT will become the main source of suicide prevention activities will depend upon its efficacy in helping people, compared with and as a complement to existing services and activities.

Although there is a growing body of evidence regarding ICT use in suicide prevention, program evaluation is still lacking. There is a need for more research evaluating and comparing the impacts of various ICT strategies in different contexts, understanding the profiles of individuals at risk of suicide who use ICT, and web-based help-seeking behaviors. We also need to better understand the impact of ICT on individuals who are bereaved by suicide. Moreover, ethics and security concerns regarding web-based suicide prevention have been the focus of very limited research and need to be addressed in future studies.

Furthermore, service users, providers, and managers from private or public systems should be informed as of now and updated regularly on the benefits and risks of ICT use in health and social services. This information should include (1) effectiveness in various well-described contexts and potential unexpected iatrogenic effects, (2) cost-benefit relationship to the best comparator, (3) access, (4) acceptability and ethical concerns, (5) security, and (6) implementation. Concerning implementation, quality standards should be similar to those of Improving Access to Psychological Therapies standards set by the United Kingdom for access to psychotherapy. These standards are (1) a model of care, (2) access, (3) evidence-based interventions, (4) outcome-based measurement, and (5) the

provider's training and supervision. International health and social technology assessment agencies should therefore consider developing guidelines and a system of voluntary accreditation for ICT use in suicide prevention. National and regional public health and social services may require that, before commissioning an ICT for suicide prevention or mental health care, their Health Technology Agency recommends ICT based on its efficacy, efficiency, safety, acceptability, and feasibility in the context of jurisdiction.

Acknowledgments

The authors would like to thank Brian Mishara, Alain Lesage, Jérôme Gaudreault, and the Association Québécoise de prevention du suicide for their valuable and constructive suggestions during the planning and development of this study. A grant was awarded for the completion of the study to RL and CB by the Canadian Institutes of Health Research, Institute of Population and Public Health (grant 385 708). JR received a postdoctoral grant from the Quebec Network on Nursing Intervention Research and the project from the Quebec Network on Suicide, Mood Disorders and Related Disorders.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy syntax specific to each consulted database.
[DOCX File , 15 KB-Multimedia Appendix 1]

References

1. Larsen ME, Nicholas J, Christensen H. A systematic assessment of smartphone tools for suicide prevention. PLoS One 2016;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]
2. van Spijker BA, Werner-Seidler A, Batterham PJ, Mackinnon A, Calear AL, Gosling JA, et al. Effectiveness of a web-based self-help program for suicidal thinking in an Australian community sample: randomized controlled trial. J Med Internet Res 2018 Feb 14;20(2):e15 [FREE Full text] [doi: 10.2196/jmir.8595] [Medline: 29444769]
3. Perry Y, Werner-Seidler A, Calear A, Christensen H. Web-based and mobile suicide prevention interventions for young people: a systematic review. J Can Acad 2016;25(2):73-79. [Medline: 27274742]
4. de la Torre I, Castillo G, Arambarri J, López-Coronado M, Franco MA. Mobile apps for suicide prevention: review of virtual stores and literature. JMIR mHealth and uHealth 2017 Oct 10;5(10):e130. [doi: 10.2196/mhealth.8036] [Medline: 29017992]
5. Robinson J, Cox G, Bailey E, Hetrick S, Rodrigues M, Fisher S, et al. Social media and suicide prevention: a systematic review. Early Interv Psychiatry 2016 Apr;10(2):103-121. [doi: 10.1111/eip.12229] [Medline: 25702826]
6. Robinson J, Bailey E, Witt K, Stefanac N, Milner A, Currier D, et al. What works in youth suicide prevention? A systematic review and meta-analysis. EClinicalMedicine 2018;4-5:52-91 [FREE Full text] [doi: 10.1016/j.eclinm.2018.10.004] [Medline: 31193651]
7. van Spijker BAJ, van Straten A, Kerkhof AJ. Effectiveness of online self-help for suicidal thoughts: results of a randomised controlled trial. PLoS ONE 2014 Feb 27;9(2):e90118. [doi: 10.1371/journal.pone.0090118]
8. Wagner B, Horn AB, Maercker A. Internet-based versus face-to-face cognitive-behavioral intervention for depression: a randomized controlled non-inferiority trial. J Affect Disord 2014 Jan;152-154:113-121. [doi: 10.1016/j.jad.2013.06.032] [Medline: 23886401]
9. Organisation mondiale de la Santé. Prévention du suicide: l'état d'urgence mondial. Geneva: Organisation mondiale de la Santé; 2014.
10. Le grand dictionnaire terminologique (GDT). Office québécois de la langue française. URL: http://gdt.oqlf.gouv.qc.ca/ [accessed 2020-09-16]
11. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
12. Levac D, Colquhoun H, O'Brien K. Scoping studies: advancing the methodology. Implementation Science 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69]
13. Mishara B, Kerkhof AJFM, editors. Suicide Prevention and New Technologies: Evidence-Based Practice. London: Palgrave MacMillan; 2013.
14. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
15. Grist R, Porter J, Stallard P. Mental Health Mobile Apps for Preadolescents and Adolescents: A Systematic Review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]

16. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, the PRISMA Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015 Jan 02;349(jan02 1):g7647-g7647 [FREE Full text] [doi: 10.1136/bmj.g7647]
17. Methely A, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. BMC Health Services Research 2014;14(1):1-10 [FREE Full text] [doi: 10.1186/s12913-014-0579-0]
18. Straus SE, Tetroe J, Graham ID, Zwarenstein M, Bhattacharyya O, Shepperd S. Monitoring use of knowledge and evaluating outcomes. Canadian Medical Association Journal 2010 Jan 18;182(2):E94-E98. [doi: 10.1503/cmaj.081335]
19. Pourmand A, Roberon J, Caggiula A, Monsalve N, Rahimi M, Torres-Llenza V. Social media and suicide: a review of technology-based epidemiology and risk assessment. Telemedicine and e-Health 2019;25(10):880-888 [FREE Full text] [doi: 10.1089/tmj.2018.0203]
20. Bartels EM. How to perform a systmatic search. Best Practice & Research Clinical Rheumatology 2013;27(2):295-306 [FREE Full text] [doi: 10.1016/j.berh.2013.02.001]
21. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
22. Mishara B, Chagnon F, Daigle M, Balan B, Raymond S, Marcoux I, et al. Comparing models of helper behavior to actual practice in telephone crisis intervention: A silent monitoring study of calls to the US 1-800-SUICIDE network. Suicide and Life-Threatening Behavior 2007;37(3):291-307. [doi: 10.1521/suli.2007.37.3.291]
23. Mishara B, Daigle M. Helplines crisis intervention services: challenges for the future. In: Lester D, editor. Suicide Prevention: resources for the millennium. New York: Routledge; 2001:153-171.
24. Mishara BL, Daigle M, Bardon C, Chagnon F, Balan B, Raymond S, et al. Comparison of the Effects of Telephone Suicide Prevention Help by Volunteers and Professional Paid Staff: Results from Studies in the USA and Quebec, Canada. Suicide Life Threat Behav 2016 Oct;46(5):577-587 [FREE Full text] [doi: 10.1111/sltb.12238] [Medline: 26946110]
25. Henschke N, Ostelo R, Vlaeven S, Morley S, Assendelft W, Main C. Behavioural treatment for chronic low-back pain. Cochrane Database System Reviews.: Cochrane Library; 2010. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7065591/pdf/CD002014.pdf [accessed 2020-09-15]
26. Schünemann H, Vist G, Higgins J, Santesso N, Deeks J, Glasziou P, et al. Interpreting results and drawing conclusions. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions, Second Edition. Chichester (UK): John Wiley & Sons; 2019:403-431.
27. Carli V. Preventing suicidality through online tools: The SUPREME Project. In: Courtet P, editor. Understanding Suicide: from Diagnosis to Personalized Treatment. New York: Springer; 2016:281-289.
28. Wexler L, Gubrium A, Griffin M, DiFulvio G. Promoting Positive Youth Development and Highlighting Reasons for Living in Northwest Alaska Through Digital Storytelling. Health Promotion Practice 2012 Oct 24;14(4):617-623. [doi: 10.1177/1524839912462390]
29. Hurley, SJ. Public pedagogy and the experience of video creators in the It Gets Better Project. Libraries digital conservancy. 2014. URL: https://hdl.handle.net/11299/171186 [accessed 2020-09-16]
30. Rogers ML, Schneider ME, Gai AR, Gorday JY, Joiner TE. Evaluation of two web-based interventions in reducing the stigma of suicide. Behaviour Research and Therapy 2018 Oct;109:49-55. [doi: 10.1016/j.brat.2018.08.001]
31. Robinson J, Bailey E, Hetrick S, Paix S, O'Donnell M, Cox G, et al. Developing Social Media-Based Suicide Prevention Messages in Partnership With Young People: Exploratory Study. JMIR Ment Health 2017 Oct 04;4(4):e40 [FREE Full text] [doi: 10.2196/mental.7847] [Medline: 28978499]
32. Till B, Tran US, Voracek M, Niederkrotenthaler T. Beneficial and harmful effects of educative suicide prevention websites: randomised controlled trial exploring Papageno . Werther effects. Br J Psychiatry 2018 Jan 02;211(2):109-115. [doi: 10.1192/bjp.bp.115.177394]
33. Taylor-Rodgers E, Batterham PJ. Evaluation of an online psychoeducation intervention to promote mental health help seeking attitudes and intentions among young adults: randomised controlled trial. J Affect Disord 2014 Oct 15;168:65-71. [doi: 10.1016/j.jad.2014.06.047] [Medline: 25038293]
34. Berger J. Adolescents on the lookout for suicidal friends on social networking sites. Digital Repository at the University of Maryland.: Doctoral dissertation, University of Maryland, College Park; 2013. URL: https://drum.lib.umd.edu/handle/1903/14494 [accessed 2020-09-22]
35. Robinson J, Rodrigues M, Fisher S, Bailey E, Herrman H. Social media and suicide prevention: findings from a stakeholder survey. Shanghai Arch Psychiatry 2015 Feb 25;27(1):27-35 [FREE Full text] [doi: 10.11919/j.issn.1002-0829.214133] [Medline: 25852253]
36. Chen JI, Smolenski DJ, Dobscha SK, Bush NE, Denneson LM. Correlates of Mental Health Smartphone Application Use among Patients with Suicidal Ideation. Journal of Technology in Human Services 2018 Nov 16;36(4):191-207. [doi: 10.1080/15228835.2018.1502119]
37. Sueki H, Ito J. Suicide prevention through online gatekeeping using search advertising techniques: a feasibility study. Crisis 2015 Jul;36(4):267-273. [doi: 10.1027/0227-5910/a000322] [Medline: 26440623]

38. Liu NH, Contreras O, Muñoz RF, Leykin Y. Assessing Suicide Attempts and Depression Among Chinese Speakers Over the Internet. Crisis 2014 Sep 01;35(5):322-329. [doi: 10.1027/0227-5910/a000261]
39. Buckingham CD, Adams A, Vail L, Kumar A, Ahmed A, Whelan A, et al. Integrating service user and practitioner expertise within a web-based system for collaborative mental-health risk and safety management. Patient Education and Counseling 2015 Oct;98(10):1189-1196. [doi: 10.1016/j.pec.2015.08.018]
40. Gryglewicz K, Chen JI, Romero GD, Karver MS, Witmeier M. Online Suicide Risk Assessment and Management Training. Crisis 2017 May;38(3):186-194. [doi: 10.1027/0227-5910/a000421]
41. Ryan K, Tindall C, Strudwick G. Enhancing Key Competencies of Health Professionals in the Assessment and Care of Adults at Risk of Suicide Through Education and Technology. Clinical Nurse Specialist 2017;31(5):268-275. [doi: 10.1097/nur.0000000000000322]
42. LeCloux M. The development of a brief suicide screening and risk assessment training webinar for rural primary care practices. Journal of Rural Mental Health 2018 Jan;42(1):60-66. [doi: 10.1037/rmh0000087]
43. Passos IC, Mwangi B, Cao B, Hamilton JE, Wu M, Zhang XY, et al. Identifying a clinical signature of suicidality among patients with mood disorders: A pilot study using a machine learning approach. J Affect Disord 2016 Mar 15;193:109-116 [FREE Full text] [doi: 10.1016/j.jad.2015.12.066] [Medline: 26773901]
44. Cook BL, Progovac AM, Chen P, Mullin B, Hou S, Baca-Garcia E. Novel Use of Natural Language Processing (NLP) to Predict Suicidal Ideation and Psychiatric Symptoms in a Text-Based Mental Health Intervention in Madrid. Computational and Mathematical Methods in Medicine 2016;2016:1-8. [doi: 10.1155/2016/8708434]
45. Pestian J, Grupp-Phelan J, Bretonnel CK, Meyers G, Richey L, Matykiewicz P, et al. A controlled trial using natural language processing to examine the language of suicidal adolescents in the emergency department. Suicide & Life-Threatening Behavior 2016;46:154-159 [FREE Full text] [doi: 10.1111/sltb.12180]
46. Pestian J, Sorter M, Connolly B, Cohen K, McCullumsmith C, Gee J, et al. A machine learning approach to identifying the thought markers of suicidal subjects: a prospective multicenter trial. Suicide and Life-Threatening Behavior 2017;47:112-121. [doi: 10.1111/sltb.12312]
47. Venek V, Scherer S, Morency L, Rizzo AS, Pestian J. Adolescent Suicidal Risk Assessment in Clinician-Patient Interaction. IEEE Trans. Affective Comput 2017 Apr 1;8(2):204-215. [doi: 10.1109/taffc.2016.2518665]
48. Braithwaite SR, Giraud-Carrier C, West J, Barnes MD, Hanson CL. Validating Machine Learning Algorithms for Twitter Data Against Established Measures of Suicidality. JMIR Ment Health 2016 May 16;3(2):e21 [FREE Full text] [doi: 10.2196/mental.4822] [Medline: 27185366]
49. Cheng Q, Li TM, Kwok C, Zhu T, Yip PS. Assessing Suicide Risk and Emotional Distress in Chinese Social Media: A Text Mining and Machine Learning Study. J Med Internet Res 2017 Jul 10;19(7):e243. [doi: 10.2196/jmir.7276]
50. Guan L, Hao B, Cheng Q, Yip PS, Zhu T. Identifying Chinese Microblog Users With High Suicide Probability Using Internet-Based Profile and Linguistic Features: Classification Model. JMIR Mental Health 2015 May 12;2(2):e17. [doi: 10.2196/mental.4227]
51. Lv M, Li A, Liu T, Zhu T. Creating a Chinese suicide dictionary for identifying suicide risk on social media. PeerJ 2015;3:e1455 [FREE Full text] [doi: 10.7717/peerj.1455] [Medline: 26713232]
52. O'Dea B, Larsen ME, Batterham PJ, Calear AL, Christensen H. A Linguistic Analysis of Suicide-Related Twitter Posts. Crisis 2017;38(5):319-329. [doi: 10.1027/0227-5910/a000443] [Medline: 28228065]
53. Ren F, Kang X, Quan C. Examining Accumulated Emotional Traits in Suicide Blogs With an Emotion Topic Model. IEEE J. Biomed. Health Inform 2016 Sep;20(5):1384-1396. [doi: 10.1109/jbhi.2015.2459683]
54. Hettige NC, Nguyen TB, Yuan C, Rajakulendran T, Baddour J, Bhagwat N, et al. Classification of suicide attempters in schizophrenia using sociocultural and clinical features: A machine learning approach. General Hospital Psychiatry 2017 Jul;47:20-28. [doi: 10.1016/j.genhosppsych.2017.03.001]
55. Batterham PJ, van Spijker BAJ, Mackinnon AJ, Calear AL, Wong Q, Christensen H. Consistency of trajectories of suicidal ideation and depression symptoms: Evidence from a randomized controlled trial. Depress Anxiety 2018 Nov 22;36(4):321-329. [doi: 10.1002/da.22863]
56. Roaten K, Johnson C, Genzel R, Khan F, North CS. Development and Implementation of a Universal Suicide Risk Screening Program in a Safety-Net Hospital System. The Joint Commission Journal on Quality and Patient Safety 2018 Jan;44(1):4-11. [doi: 10.1016/j.jcjq.2017.07.006]
57. Adamou M, Antoniou G, Greasidou E, Lagani V, Charonyktakis P, Tsamardinos I, et al. Toward Automatic Risk Assessment to Support Suicide Prevention. Crisis 2019 Jul;40(4):249-256. [doi: 10.1027/0227-5910/a000561]
58. Silvén Hagström A. 'Suicide stigma' renegotiated: Storytelling, social support and resistance in an Internet-based community for the young suicide-bereaved. Qualitative Social Work 2016 Apr 27;16(6):775-792. [doi: 10.1177/1473325016644039]
59. Burnap P, Colombo G, Amery R, Hodorog A, Scourfield J. Multi-class machine classification of suicide-related communication on Twitter. Online Soc Netw Media 2017;2:32-44 [FREE Full text] [doi: 10.1016/j.osnem.2017.08.001]
60. Du J, Zhang Y, Luo J, Jia Y, Wei Q, Tao C, et al. Extracting psychiatric stressors for suicide from social media using deep learning. BMC Med Inform Decis Mak(Suppl 2) 2018;18:43 [FREE Full text] [doi: 10.1186/s12911-018-0632-8]
61. Coppersmith G, Leary R, Crutchley P, Fine A. Natural Language Processing of Social Media as Screening for Suicide Risk. Biomed Inform Insights 2018 Aug 27;10:117822261879286. [doi: 10.1177/1178222618792860]

62. Jung H, Park H, Song T. Ontology-Based Approach to Social Data Sentiment Analysis: Detection of Adolescent Depression Signals. J Med Internet Res 2017 Jul 24;19(7):e259. [doi: 10.2196/jmir.7452]
63. Silverstone PH, Bercov M, Suen VYM, Allen A, Cribben I, Goodrick J, et al. Long-term Results from the Empowering a Multimodal Pathway Toward Healthy Youth Program, a Multimodal School-Based Approach, Show Marked Reductions in Suicidality, Depression, and Anxiety in 6,227 Students in Grades 6–12 (Aged 11–18). Front. Psychiatry 2017 May 15;8:81. [doi: 10.3389/fpsyt.2017.00081]
64. Downs N, Feng W, Kirby B, McGuire T, Moutier C, Norcross W, et al. Listening to Depression and Suicide Risk in Medical Students: the Healer Education Assessment and Referral (HEAR) Program. Acad Psychiatry 2014 Apr 5;38(5):547-553. [doi: 10.1007/s40596-014-0115-x]
65. Haskins J, Carson JG, Chang CH, Kirshnit C, Link DP, Navarra L, et al. The Suicide Prevention, Depression Awareness, and Clinical Engagement Program for Faculty and Residents at the University of California, Davis Health System. Acad Psychiatry 2015 Jun 11;40(1):23-29. [doi: 10.1007/s40596-015-0359-0]
66. Williams D. Consensus-based strategies for using social networking to reduce suicide risk: a Delphi study. In: Doctoral Dissertation. Denver: The University of the Rockies; 2016.
67. King CA, Eisenberg D, Zheng K, Czyz E, Kramer A, Horwitz A, et al. Online suicide risk screening and intervention with college students: a pilot randomized controlled trial. J Consult Clin Psychol 2015 Jun;83(3):630-636 [FREE Full text] [doi: 10.1037/a0038805] [Medline: 25688811]
68. Silverstone PH, Bercov M, Suen VYM, Allen A, Cribben I, Goodrick J, et al. Initial Findings from a Novel School-Based Program, EMPATHY, Which May Help Reduce Depression and Suicidality in Youth. PLoS ONE 2015 May 14;10(5):e0125527. [doi: 10.1371/journal.pone.0125527]
69. Davidson JE, Zisook S, Kirby B, DeMichele G, Norcross W. Suicide Prevention: a healer education and referral program for nurses. JONA: The Journal of Nursing Administration 2018;48(2):85-92. [doi: 10.1097/nna.0000000000000582]
70. Goodyear-Smith F, Martel R, Darragh M, Warren J, Thabrew H, Clark TC. Screening for risky behaviour and mental health in young people: the YouthCHAT programme. Public Health Rev 2017 Oct;38:20 [FREE Full text] [doi: 10.1186/s40985-017-0068-1] [Medline: 29450092]
71. Christensen H, Farrer L, Batterham PJ, Mackinnon A, Griffiths KM, Donker T. The effect of a web-based depression intervention on suicide ideation: secondary outcome from a randomised controlled trial in a helpline. BMJ Open 2013 Jun 28;3(6):e002886. [doi: 10.1136/bmjopen-2013-002886]
72. Mewton L, Andrews G. Cognitive behaviour therapy via the internet for depression: a useful strategy to reduce suicidal ideation. J Affect Disord 2015 Jan 01;170:78-84. [doi: 10.1016/j.jad.2014.08.038] [Medline: 25233243]
73. Whiteside U, Richards J, Steinfeld B, Simon G, Caka S, Tachibana C, et al. Online cognitive behavioral therapy for depressed primary care patients: a pilot feasibility project. Perm J 2014 May;18(2):21-27 [FREE Full text] [doi: 10.7812/TPP/13-155] [Medline: 24867546]
74. Saulsberry A, Marko-Holguin M, Blomeke K, Hinkle C, Fogel J, Gladstone T, et al. Randomized Clinical Trial of a Primary Care Internet-based Intervention to Prevent Adolescent Depression: One-year Outcomes. J Can Acad Child Adolesc Psychiatry 2013 May;22(2):106-117 [FREE Full text] [Medline: 23667356]
75. Guille C, Zhao Z, Krystal J, Nichols B, Brady K, Sen S. Web-Based Cognitive Behavioral Therapy Intervention for the Prevention of Suicidal Ideation in Medical Interns: A Randomized Clinical Trial. JAMA Psychiatry 2015 Dec;72(12):1192-1198 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.1880] [Medline: 26535958]
76. Hedman E, Ljótsson B, Kaldo V, Hesser H, El Alaoui S, Kraepelien M, et al. Effectiveness of Internet-based cognitive behaviour therapy for depression in routine psychiatric care. Journal of Affective Disorders 2014 Feb;155:49-58. [doi: 10.1016/j.jad.2013.10.023]
77. Nielssen O, Dear BF, Staples LG, Dear R, Ryan K, Purtell C, et al. Procedures for risk management and a review of crisis referrals from the MindSpot Clinic, a national service for the remote assessment and treatment of anxiety and depression. BMC Psychiatry 2015;15:304 [FREE Full text] [doi: 10.1186/s12888-015-0676-6] [Medline: 26626712]
78. Williams AD, Andrews G. The Effectiveness of Internet Cognitive Behavioural Therapy (iCBT) for Depression in Primary Care: A Quality Assurance Study. PLoS ONE 2013 Feb 22;8(2):e57447. [doi: 10.1371/journal.pone.0057447]
79. Lara MA, Tiburcio M, Aguilar AA, Sánchez-Solís A. A four-year experience with a Web-based self-help intervention for depressive symptoms in Mexico. Rev Panam Salud Publica 2014;35(5-6):399-406 [FREE Full text] [Medline: 25211568]
80. Gleeson J, Lederman R, Wadley G, Bendall S, McGorry P, Alvarez-Jimenez M. Safety and privacy outcomes from a moderated online social therapy for young people with first-episode psychosis. Psychiatric services 2014;65(4):546-550 [FREE Full text] [doi: 10.1176/appi.ps.201300078]
81. Whiteside U, Lungu A, Richards J, Simon GE, Clingan S, Siler J, et al. Designing Messaging to Engage Patients in an Online Suicide Prevention Intervention: Survey Results From Patients With Current Suicidal Ideation. J Med Internet Res 2014 Feb 07;16(2):e42. [doi: 10.2196/jmir.3173]
82. Ghoncheh R, Gould MS, Twisk JW, Kerkhof AJ, Koot HM. Efficacy of Adolescent Suicide Prevention E-Learning Modules for Gatekeepers: A Randomized Controlled Trial. JMIR Ment Health 2016 Jan 29;3(1):e8 [FREE Full text] [doi: 10.2196/mental.4614] [Medline: 26825006]

83. Lamis DA, Underwood M, D'Amore N. Outcomes of a Suicide Prevention Gatekeeper Training Program Among School Personnel. Crisis 2017 Mar;38(2):89-99. [doi: 10.1027/0227-5910/a000414]
84. Smith AR, Silva C, Covington DW, Joiner TE. An assessment of suicide-related knowledge and skills among health professionals. Health Psychology 2014;33(2):110-119. [doi: 10.1037/a0031062]
85. Marshall E, York J, Magruder K, Yeager D, Knapp R, De Santis ML, et al. Implementation of Online Suicide-Specific Training for VA Providers. Acad Psychiatry 2014 Feb 22;38(5):566-574. [doi: 10.1007/s40596-014-0039-5]
86. Krieg C. Can demonstration enhance the effects of an online risk assessment training workshop? Doctoral dissertation, Counseling Psychology. Arizona State University. ASU Library. Digital Repository. 2017. URL: https://repository.asu.edu/items/40196 [accessed 2020-09-17]
87. Rigsbee N. Suicide assessment training in counselor education. Doctoral dissertation. University of New Mexico. UNM Digital Repository. Albuquerque, New Mexico: The University of New Mexico; 2015. URL: https://tinyurl.com/3vzfr7bs [accessed 2020-09-17]
88. Robinson J, Hetrick S, Cox G, Bendall S, Yuen HP, Yung A, et al. Can an Internet-based intervention reduce suicidal ideation, depression and hopelessness among secondary school students: results from a pilot study. Early Interv Psychiatry 2016 Feb;10(1):28-35. [doi: 10.1111/eip.12137] [Medline: 24684946]
89. Skovgaard Larsen JL, Frandsen H, Erlangsen A. MYPLAN – A Mobile Phone Application for Supporting People at Risk of Suicide. Crisis 2016 May;37(3):236-240. [doi: 10.1027/0227-5910/a000371]
90. Rizvi S, Hughes C, Thomas M. The DBT Coach mobile application as an adjunct to treatment for suicidal and self-injuring individuals with borderline personality disorder: a preliminary evaluation and challenges to client utilization. Psychological Services 2016;13(4):380-388. [doi: 10.1037/ser0000100]
91. Larsen ME, Shand F, Morley K, Batterham PJ, Petrie K, Reda B, et al. A Mobile Text Message Intervention to Reduce Repeat Suicidal Episodes: Design and Development of Reconnecting After a Suicide Attempt (RAFT). JMIR Ment Health 2017 Dec 13;4(4):e56. [doi: 10.2196/mental.7500]
92. AL-Asadi AM, Klein B, Meyer D. Multiple Comorbidities of 21 Psychological Disorders and Relationships With Psychosocial Variables: A Study of the Online Assessment and Diagnostic System Within a Web-Based Population. J Med Internet Res 2015 Feb 26;17(3):e55. [doi: 10.2196/jmir.4143]
93. Hesdorffer DC, French JA, Posner K, DiVentura B, Pollard JR, Sperling MR, et al. Suicidal ideation and behavior screening in intractable focal epilepsy eligible for drug trials. Epilepsia 2013 Feb 28;54(5):879-887. [doi: 10.1111/epi.12128]
94. Best P, Foye U, Taylor B, Hazlett D, Manktelow R. Online interactive suicide support services: quality and accessibility. Mental Health Review Journal 2013 Nov 29;18(4):226-239. [doi: 10.1108/mhrj-03-2013-0009]
95. Flamarique I, Santosh P, Zuddas A, Arango C, Purper-Ouakil D, Hoekstra P, STOP consortium. Development and psychometric properties of the Suicidality: Treatment Occurring in Paediatrics (STOP) Suicidality Assessment Scale (STOP-SAS) in children and adolescents. BMC pediatrics 2016;16(1):213 [FREE Full text] [doi: 10.1186/s12887-016-0751-2]
96. Hirschhorn E. Integrating technology into collaborative suicide risk assessment: comparing electronic and paper-and-pencil versions of the suicide status form. In: Doctoral dissertation. Washington D.C: Catholic University of America; 2017.
97. AL-Asadi AM, Klein B, Meyer D. Comorbidity Structure of Psychological Disorders in the Online e-PASS Data as Predictors of Psychosocial Adjustment Measures: Psychological Distress, Adequate Social Support, Self-Confidence, Quality of Life, and Suicidal Ideation. J Med Internet Res 2014 Oct 28;16(10):e248. [doi: 10.2196/jmir.3591]
98. De Beurs DP, de Vries AL, de Groot MH, de Keijser J, Kerkhof AJ. Applying Computer Adaptive Testing to Optimize Online Assessment of Suicidal Behavior: A Simulation Study. J Med Internet Res 2014 Sep 11;16(9):e207. [doi: 10.2196/jmir.3511]
99. Iorfino F, Davenport TA, Ospina-Pinillos L, Hermens DF, Cross S, Burns J, et al. Using New and Emerging Technologies to Identify and Respond to Suicidality Among Help-Seeking Young People: A Cross-Sectional Study. J Med Internet Res 2017 Jul 12;19(7):e247. [doi: 10.2196/jmir.7897]
100. Haner D, Pepler D. "Live Chat" Clients at Kids Help Phone: Individual Characteristics and Problem Topics. J Can Acad Child Adolesc Psychiatry 2016;25(3):138-144 [FREE Full text] [Medline: 27924143]
101. Bennett A, Pourmand A, Shokoohi H, Shesser R, Sanchez J, Joyce J. Impacts of Social Networking Sites on Patient Care in the Emergency Department. Telemedicine and e-Health 2014 Jan;20(1):94-96. [doi: 10.1089/tmj.2013.0055]
102. Batterham PJ, Calear AL, Farrer L, McCallum SM, Cheng VWS. FitMindKit : Randomised controlled trial of an automatically tailored online program for mood, anxiety, substance use and suicidality. Internet Interventions 2018 Jun;12:91-99. [doi: 10.1016/j.invent.2017.08.002]
103. Rice S, Robinson J, Bendall S, Hetrick S, Cox G, Bailey E, et al. Online and Social Media Suicide Prevention Interventions for Young People: A Focus on Implementation and Moderation. J Can Acad Child Adolesc Psychiatry 2016;25(2):80-86 [FREE Full text] [Medline: 27274743]
104. Bresó A, Martínez-Miranda J, Botella C, Baños R, García-Gómez J. Usability and acceptability assessment of an empathic virtual agent to prevent major depression. Expert Systems: International Journal of Knowledge Engineering and Neural Networks 2016;33(4):297-312. [doi: 10.1111/exsy.12151]

105. Madsen T, van Spijker B, Karstoft K, Nordentoft M, Kerkhof AJ. Trajectories of Suicidal Ideation in People Seeking Web-Based Help for Suicidality: Secondary Analysis of a Dutch Randomized Controlled Trial. J Med Internet Res 2016 Jun 30;18(6):e178. [doi: 10.2196/jmir.5904]
106. Hetrick SE, Goodall J, Yuen HP, Davey CG, Parker AG, Robinson J, et al. Comprehensive Online Self-Monitoring to Support Clinicians Manage Risk of Suicide in Youth Depression. Crisis 2017 May;38(3):147-157. [doi: 10.1027/0227-5910/a000422]
107. Hetrick SE, Yuen HP, Bailey E, Cox GR, Templer K, Rice SM, et al. Internet-based cognitive behavioural therapy for young people with suicide-related behaviour (Reframe-IT): a randomised controlled trial. Evid Based Mental Health 2017 Jul 12;20(3):76-82. [doi: 10.1136/eb-2017-102719]
108. Batterham PJ, Christensen H, Mackinnon AJ, Gosling JA, Thorndike FP, Ritterband LM, et al. Trajectories of change and long-term outcomes in a randomised controlled trial of internet-based insomnia treatment to prevent depression. BJPsych open 2018 Jan 02;3(5):228-235. [doi: 10.1192/bjpo.bp.117.005231]
109. Wilksch SM, O'Shea A, Wade TD. Depressive symptoms, alcohol and other drug use, and suicide risk: Prevention and treatment effects from a two-country online eating disorder risk reduction trial. Int J Eat Disord 2018 Dec 23;52(2):132-141. [doi: 10.1002/eat.23005]
110. Maulik PK, Kallakuri S, Devarapalli S. Operational challenges in conducting a community-based technology-enabled mental health services delivery model for rural India: Experiences from the SMART Mental Health Project. Wellcome Open Res 2018 Apr 18;3:43. [doi: 10.12688/wellcomeopenres.14524.1]
111. Reins JA, Boß L, Lehr D, Berking M, Ebert DD. The more I got, the less I need? Efficacy of Internet-based guided self-help compared to online psychoeducation for major depressive disorder. Journal of Affective Disorders 2019 Mar;246:695-705. [doi: 10.1016/j.jad.2018.12.065]
112. Richards D, Duffy D, Burke J, Anderson M, Connell S, Timulak L. Supported Internet-Delivered Cognitive Behavior Treatment for Adults with Severe Depressive Symptoms: A Secondary Analysis. JMIR Ment Health 2018 Oct 02;5(4):e10204. [doi: 10.2196/10204]
113. Tewari A, Kallakuri S, Devarapalli S, Jha V, Patel A, Maulik PK. Process evaluation of the systematic medical appraisal, referral and treatment (SMART) mental health project in rural India. BMC Psychiatry 2017 Dec 4;17(1):385. [doi: 10.1186/s12888-017-1525-6]
114. Wilks CR, Lungu A, Ang SY, Matsumiya B, Yin Q, Linehan MM. A randomized controlled trial of an Internet delivered dialectical behavior therapy skills training for suicidal and heavy episodic drinkers. Journal of Affective Disorders 2018 May;232:219-228. [doi: 10.1016/j.jad.2018.02.053]
115. Labelle R, Bibaud-De SA, Leblanc F. Innovating to treat depressionprevent suicide: the IPhone@ PSY ASSISTANCE application. In: Mishara B, Kerkhof AJFM, editors. Suicide Prevention and New Technologies: Evidence-Based Practices. London: Palgrave Macmillan; 2013:166-180.
116. Bush NE, Dobscha SK, Crumpton R, Denneson LM, Hoffman JE, Crain A, et al. A Virtual Hope Box smartphone app as an accessory to therapy: proof-of-concept in a clinical sample of veterans. Suicide Life Threat Behav 2015 Feb;45(1):1-9. [doi: 10.1111/sltb.12103] [Medline: 24828126]
117. Povey J, Mills PPJR, Dingwall KM, Lowell A, Singer J, Rotumah D, et al. Acceptability of Mental Health Apps for Aboriginal and Torres Strait Islander Australians: A Qualitative Study. J Med Internet Res 2016;18(3):e65 [FREE Full text] [doi: 10.2196/jmir.5314] [Medline: 26969043]
118. Mackie C, Dunn N, MacLean S, Testa V, Heisel M, Hatcher S. A qualitative study of a blended therapy using problem solving therapy with a customised smartphone app in men who present to hospital with intentional self-harm. Evid Based Mental Health 2017 Oct 13;20(4):118-122. [doi: 10.1136/eb-2017-102764]
119. Mok K, Jorm AF, Pirkis J. Who Goes Online for Suicide-Related Reasons? Crisis 2016 Mar;37(2):112-120. [doi: 10.1027/0227-5910/a000366]
120. Hetrick SE, Robinson J, Burge E, Blandon R, Mobilio B, Rice SM, et al. Youth Codesign of a Mobile Phone App to Facilitate Self-Monitoring and Management of Mood Symptoms in Young People With Major Depression, Suicidal Ideation, and Self-Harm. JMIR Ment Health 2018 Jan 23;5(1):e9 [FREE Full text] [doi: 10.2196/mental.9041] [Medline: 29362208]
121. O'Toole M, Arendt M, Pedersen C. Testing an app-assisted treatment for suicide prevention in a randomized controlled trial: effects on suicide risk and depression. Behavior Therapy 2019;50(2):421-429 [FREE Full text] [doi: 10.1016/j.beth.2018.07.007]
122. Hetrick SE, Dellosa MK, Simmons MB, Phillips L. Development and pilot testing of an online monitoring tool of depression symptoms and side effects for young people being treated for depression. Early Intervention in Psychiatry 2014 Feb 25;9(1):66-69. [doi: 10.1111/eip.12127]
123. Chen RY, Feltes JR, Tzeng WS, Lu ZY, Pan M, Zhao N, et al. Phone-Based Interventions in Adolescent Psychiatry: A Perspective and Proof of Concept Pilot Study With a Focus on Depression and Autism. JMIR Res Protoc 2017 Jun 16;6(6):e114. [doi: 10.2196/resprot.7245]
124. Berrouiguet S, Gravey M, Le Galudec M, Alavi Z, Walter M. Post-acute crisis text messaging outreach for suicide prevention: A pilot study. Psychiatry Research 2014 Jul;217(3):154-157. [doi: 10.1016/j.psychres.2014.02.034]

125. Kodama T, Syouji H, Takaki S, Fujimoto H, Ishikawa S, Fukutake M, et al. Text messaging for psychiatric outpatients: effect on help-seeking and self-harming behaviors. Journal of Psychosocial Nursing and Mental Health Services 2016;54(4):31-37 [FREE Full text] [doi: 10.3928/02793695-20160121-01]
126. Owens C, Charles N. Implementation of a text-messaging intervention for adolescents who self-harm (TeenTEXT): a feasibility study using normalisation process theory. Child and Adolescent Psychiatry and Mental Health 2016;10(1):14 [FREE Full text] [doi: 10.1186/s13034-016-0101-z]
127. Predmore Z, Ramchand R, Ayer L, Kotzias V, Engel C, Ebener P, et al. Expanding Suicide Crisis Services to Text and Chat. Crisis 2017 Jul;38(4):255-260. [doi: 10.1027/0227-5910/a000460]
128. Mokkenstorm JK, Eikelenboom M, Huisman A, Wiebenga J, Gilissen R, Kerkhof AJFM, et al. Evaluation of the 113Online Suicide Prevention Crisis Chat Service: Outcomes, Helper Behaviors and Comparison to Telephone Hotlines. Suicide Life Threat Behav 2016 Aug 19;47(3):282-296. [doi: 10.1111/sltb.12286]
129. Ahuja AK, Biesaga K, Sudak DM, Draper J, Womble A. Suicide on facebook. J Psychiatr Pract 2014 Mar;20(2):141-146. [doi: 10.1097/01.pra.0000445249.38801.d1] [Medline: 24638049]
130. Reyes-Portillo JA, Chin EM, Toso-Salman J, Blake Turner J, Vawdrey D, Mufson L. Using Electronic Health Record Alerts to Increase Safety Planning with Youth At-Risk for Suicide: A Non-randomized Trial. Child Youth Care Forum 2018 Jan 24;47(3):391-402. [doi: 10.1007/s10566-018-9435-4]
131. Kennard BD, Biernesser C, Wolfe KL, Foxwell AA, Craddock Lee SJ, Rial KV, et al. Developing a Brief Suicide Prevention Intervention and Mobile Phone Application: A Qualitative Report. Journal of Technology in Human Services 2015 Dec 14;33(4):345-357. [doi: 10.1080/15228835.2015.1106384]
132. Tan Z, Liu X, Liu X, Cheng Q, Zhu T. Designing Microblog Direct Messages to Engage Social Media Users With Suicide Ideation: Interview and Survey Study on Weibo. J Med Internet Res 2017 Dec 12;19(12):e381. [doi: 10.2196/jmir.8729]
133. LeMay R. Student learning of violence prevention education concepts: a longitudinal analysis. Doctoral Dissertation. 2015. URL: https://www.semanticscholar.org/paper/Student-Learning-of-Violence-Prevention-Education-A-LeMay/5217862648ffdff4702a835be8497d502f237393#citing-papers [accessed 2020-09-17]
134. de Andrade NNG, Pawson D, Muriello D, Donahue L, Guadagno J. Ethics and artificial intelligenceuicide prevention on Facebook. Philosophy & Technology 2018;31(4):669-684. [doi: 10.1007/s13347-018-0336-0]
135. McKernan LC, Clayton EW, Walsh CG. Protecting Life While Preserving Liberty: Ethical Recommendations for Suicide Prevention With Artificial Intelligence. Front. Psychiatry 2018 Dec 3;9:650. [doi: 10.3389/fpsyt.2018.00650]

Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
HEAR: Health Education Assessment and Referral
iCBT: internet-based cognitive behavioral therapy
ICT: information and communication technology
PICO: population, intervention, comparison, and outcomes
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by R Kukafka; submitted 26.10.20; peer-reviewed by N Maltais, E Bailey, K Baudin; comments to author 16.12.20; revised version received 10.02.21; accepted 16.03.21; published 04.05.21

Please cite as:
Rassy J, Bardon C, Dargis L, Côté LP, Corthésy-Blondin L, Mörch CM, Labelle R
Information and Communication Technology Use in Suicide Prevention: Scoping Review
J Med Internet Res 2021;23(5):e25288
URL: https://www.jmir.org/2021/5/e25288
doi: 10.2196/25288
PMID: 33820754