JMIR Medical Informatics

Impact Factor (2023): 3.1
Volume 5 (2017), Issue 2 ISSN 2291-9694 Editor in Chief: Christian Lovis, MD, MPH, FACMI

Contents

Original Papers

Original Paper

Telemedicine Services for the Arctic: A Systematic Review

Ashenafi Zebene Woldaregay[1], MSc (Computer Engineering), MSc (Telemedicine and e-Health); Ståle Walderhaug[1,2], PhD; Gunnar Hartvigsen[1,3], PhD, Professor

[1]Department of Computer Science, University of Tromsø - The Arctic University of Norway, Tromsø, Norway

[2]SINTEF Digital, Software Engineering, Safety and Security, Tromsø, Norway

[3]Norwegian Centre for e-Health Research, University Hospital of North Norway, Tromsø, Norway

Corresponding Author:
Ashenafi Zebene Woldaregay, MSc (Computer Engineering), MSc (Telemedicine and e-Health)
Department of Computer Science
University of Tromsø - The Arctic University of Norway
Realfagbygget Hansine Hansens vei 54 Breivika
Tromsø, 9019
Norway
Phone: 47 48682581
Fax: 47 776 44580
Email: ashenafi.z.woldaregay@uit.no

Abstract

Background: Telemedicine services have been successfully used in areas where there are adequate infrastructures such as reliable power and communication lines. However, despite the increasing number of merchants and seafarers, maritime and Arctic telemedicine have had limited success. This might be linked with various factors such as lack of good infrastructure, lack of trained onboard personnel, lack of Arctic-enhanced telemedicine equipment, extreme weather conditions, remoteness, and other geographical challenges.

Objective: The purpose of this review was to assess and analyze the current status of telemedicine services in the context of maritime conditions, extreme weather (ie, Arctic weather), and remote accidents and emergencies. Moreover, the paper aimed to identify successfully implemented telemedicine services in the Arctic region and in maritime settings and remote emergency situations and present state of the art systems for these areas. Finally, we identified the status quo of telemedicine services in the context of search and rescue (SAR) scenarios in these extreme conditions.

Methods: A rigorous literature search was conducted between September 7 and October 28, 2015, through various online databases. Peer reviewed journals and articles were considered. Relevant articles were first identified by reviewing the title, keywords, and abstract for a preliminary filter with our selection criteria, and then we reviewed full-text articles that seemed relevant. Information from the selected literature was extracted based on some predefined categories, which were defined based on previous research and further elaborated upon via iterative brainstorming.

Results: The initial hits were vetted using the title, abstract, and keywords, and we retrieved a total of 471 papers. After removing duplicates from the list, 422 records remained. Then, we did an independent assessment of the articles and screening based on the inclusion and exclusion criteria, which eliminated another 219 papers, leaving 203 relevant papers. After a full-text assessment, 36 articles were left, which were critically analyzed. The inter-rater agreement was measured using Cohen Kappa test, and disagreements were resolved through discussion.

Conclusions: Despite the increasing number of fishermen and other seafarers, Arctic and maritime working conditions are mainly characterized by an absence of access to health care facilities. The condition is further aggravated for fishermen and seafarers who are working in the Arctic regions. In spite of the existing barriers and challenges, some telemedicine services have recently been successfully delivered in these areas. These services include teleconsultation (9/37, 24%), teleradiology (8/37, 22%), teledermatology and tele-education (3/37, 8%), telemonitoring and telecardiology (telesonography) (1/37, 3%), and others (10/37, 27%). However, the use of telemedicine in relation to search and rescue (SAR) services is not yet fully exploited. Therefore, we foresee that these implemented and evaluated telemedicine services will serve as underlying models for the successful implementation of future search and rescue (SAR) services.

(JMIR Med Inform 2017;5(2):e16) doi:10.2196/medinform.6323

KEYWORDS

telemedicine; telehealth; health services accessibility; extreme cold; arctic regions; accidents

Introduction

The advent of information and communication technology (ICT) has revolutionized various aspects of health care services, including the practice of telemedicine. Nowadays, telemedicine services have a great impact on making a patient's life easier by allowing remote diagnosis and treatment without the constraints of distance and time. Many groups have performed and evaluated different clinical trials onshore and reported the success of telemedicine services [1-5]. However, these success stories are limited to onshore situations, where there are adequate infrastructures. In contrast, maritime telemedicine has received little attention despite the increasing number of seafarers from various nations around the world [6,7]. For example, Duffy's review [8] regarding the offshore dental problem of Shell Expro indicates that most of the dental problems that resulted in evacuation were preventable while on-board. This study clearly shows that a considerable number of offshore workers suffer from untreated dental diseases, resulting in unnecessary evacuations with associated costs [8]. Moreover, Guitton [7] has assessed the availability of online resources associated with maritime health. The findings show that there is a lack of such resources, and when they exist, they suffer from poor content in terms of understandability and actionability by general audiences, when compared with other medical scenarios. Adoption of onshore technology and research results in offshore conditions might seem like a quick remedy for the case. However, this requires careful investigation of each service so as to identify the underlying difference between the onshore and offshore scenarios. According to Guitton [9], maritime and onshore telemedicine can be convergent and divergent with respect to differences in structure, practices, and policy. Structural differences mainly include differences in data transfer capability (eg, onshore: easy due to availability of enough bandwidth; offshore: difficult due to absence of enough bandwidth), levels of literacy (eg, onshore: mostly between health professionals; offshore: between laymen (seafarers) or paramedics and professionals), supplementary medical assistance (eg, onshore: limited time delays before assistance arrives; offshore: very long time delays before assistance arrives). Practical differences include health records (eg, onshore: complete health records; offshore: incomplete health records), language (eg, onshore: mostly patient's mother tongue; offshore: potentially not and typically multilingual because of crossing of multiple boundaries), legal context (eg, onshore: typically deployed within a single nation; offshore: multiple nations may be involved), and complementary medical exams (eg, onshore: many; offshore: almost impossible). Policy differences mainly include policy dynamics (eg, onshore: relatively new and fast growing; offshore: old and stable), and research target (eg, onshore: targets of major research, offshore: very few) [9]. Therefore, it is necessary to identify these differences and carefully review them before transferring technology, research results, and experiences to offshore conditions [6,9]. Moreover, Pedersen et al [5] highlighted the different factors, namely, access, quality of care, cost effectiveness, and emergency care, for transferring knowledge from onshore to offshore, by justifying the case with the medical application tested and implemented on an operative basis in Northern Norway. Despite these differences, it is important to note that there is an area where direct adoption of onshore telemedicine services can be fruitful, such as patient-targeted telemedicine interventions, radio-consultation equipped with pictures and video, and video conferencing [9]. Furthermore, research results and experiences from within a similar environment, such as extreme weather conditions like those in Antarctica, can be one area to investigate when developing telemedicine solutions for the Arctic [4].

Despite the increasing number of fishermen and other seafarers, maritime working conditions are characterized by an absence of access to health care facilities [6,7]. There are factors that affect successful implementation of maritime or offshore telemedicine in the Arctic, including long distance, extreme weather conditions, absence of good communication coverage, and the time required for search and rescue (SAR) helicopters to reach the Artic, which reduce the possibility of medical evacuation (MEDEVAC) [6,10-12]. According to AH Gundersen, Senior adviser at The Joint Rescue Coordination Centre of Northern Norway (JRCC NN) in Bodø, time is considered to be the scarcest resource in an emergency situation, particularly in the Arctic, because of the long distance and harsh environmental conditions [13]. Moreover, the absence of on-board trained nurses or physicians, limited equipment and medicine, and onshore professional advice limited to only radio medical advice, further aggravates the situation [6]. Nowadays, there is a growing interest in the Arctic region in connection with the discovery of huge gas and oil resources [14,15]. However, it remains to be solved how appropriate health care services can be offered in this area. Remoteness, Arctic winter darkness, extreme weather conditions, and poor communication network coverage in the area pose a challenge for the use of remote telemedicine services [4,10,15-18]. Moreover, during a large-scale accident such as a shipwreck, the temperature of the region, which might go below −40^{o} C [16,19,20], further reduces the chances of victims surviving until search and rescue teams arrive. This situation has posed additional challenges for the implementation of successful emergency telemedicine services in the Arctic region.

Generally speaking, a review of telemedicine services in the context of maritime conditions, remote and extreme weather setting are inadequate. However, there are several reviews on telemedicine services regarding onshore remote accident and emergency services that have been published [21-24]. For example, Keane [22] conducted a comprehensive review on the success of telemedicine in the scope of accident and emergency. Amadi et al [23] also conducted a review to examine the history and existing applications of telemedicine in pre-hospital environments, where telemedicine is believed to extend the reach of specialist services to handle pre-hospital care of acute emergencies, in cases where treatment delays may affect the clinical outcomes. The purpose of this review is to assess and

analyze the status of telemedicine services, focusing on services that can fit the diverse nature of the Arctic region (environment), which is offshore, remote, and has extreme weather conditions. Moreover, it presents state of the art systems of implemented telemedicine services in remote Arctic regions and also analyses these services within the context of search and rescue services (SAR).

Methods

For the purpose of the study, we conducted a rigorous literature search between September 7 and October 28, 2015, through various online databases. The searched databases included Google Scholar, PubMed (Medline), Science Direct, ACM Digital Library, IEEE Xplore, Onepetro, the Journal of American Medical Informatics Association (JAMIA), the Journal of International Maritime Health and the Journal of Telemedicine and Telecare. Furthermore, additional articles are also extracted from the reference lists of the selected papers in order to get a complete overview of the state of the art systems. Peer-reviewed journals and articles published between 1995 and 2015 were considered. The inclusion and exclusion criteria were setup through rigorous discussion and brainstorming among the authors. Several combinations of the terms “Arctic,” “oil and gas,” “shipping,” “telemedicine,” “search and rescue,” “maritime medicine,” “offshore,” “extreme weather,” and “telehealth” were used during the search. The search strings were combined using “AND” and “OR” for a better searching strategy. The titles, keywords and abstracts were used for a preliminary filter with our selection criteria to identify relevant articles, we then reviewed full texts for articles that seemed relevant. Some predefined categories were used for information extraction from the selected literatures, which were defined based on previous research and also further elaborated upon via iterative brainstorming. Please note that the words “telemedicine” and “telehealth” are used interchangeably throughout our discussion.

Inclusion and Exclusion Criteria

To be included in the review, the studies had to have a direct involvement of telemedicine and eHealth in the following scenarios: maritime or offshore conditions and shipping, oil and gas, search and rescue, Arctic and extreme weather conditions, and remote accidents and emergencies. The studies were expected to describe solutions and implement and evaluate telemedicine services to be included in the review. Therefore, studies that were outside of the scope mentioned above were excluded from the review. Moreover, studies in all other language but English are excluded from the review.

Data Categorization and Data Collection

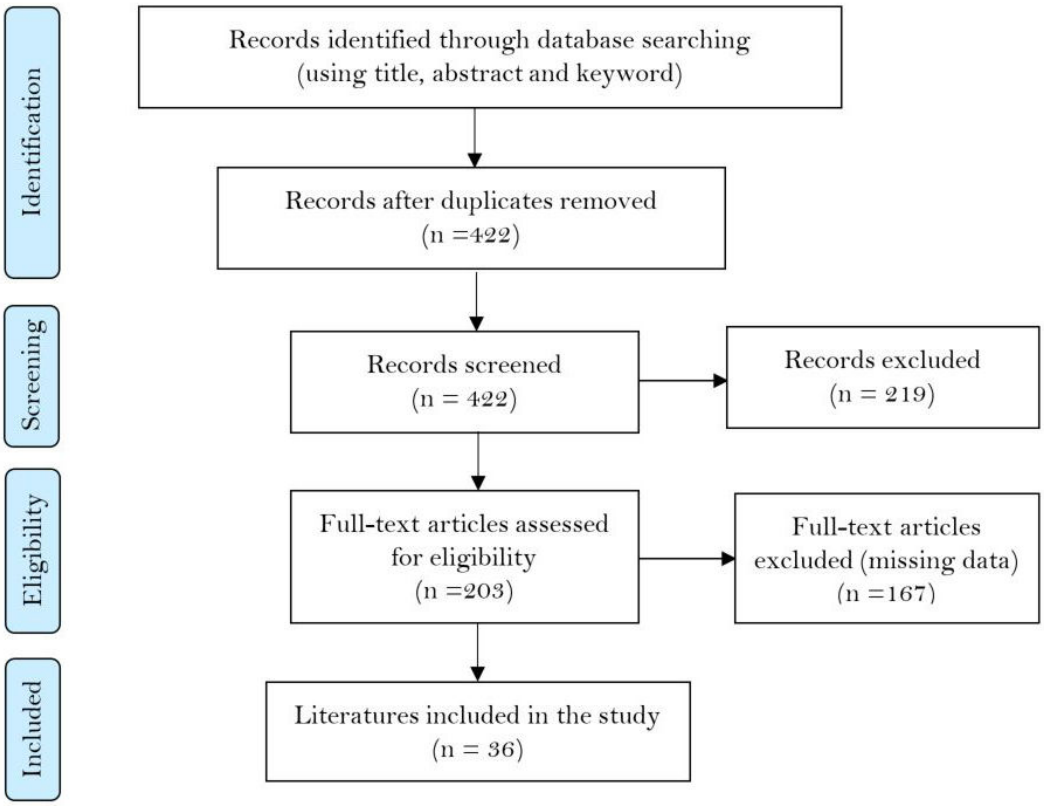
Information was extracted from the literature based on defined categories (variables). These categories were defined based on previous research, literature reviews, and also further elaborated upon via iterative brainstorming. The categories included were solely defined to assess, analyze, and evaluate the current status of telemedicine services in maritime or offshore, remote accident and emergency, and extreme weather (Arctic, Antarctica) settings. The categories we decided to include were as follows:

- Type of Circumstance: This category defines the circumstances in which the telemedicine services were delivered during the study period, that is, maritime or offshore, remote accident and emergency, or extreme weather (Arctic, Antarctica) conditions.
- Communication Link: This category defines the communication link used to facilitate the practice of telemedicine in these circumstances. This includes various communication networks such as satellite, radio, mobile, dial-up, DSL, and broadband.
- Telemedicine Modalities: This category defines the underlying modalities that enable telemedicine services to be provided. It includes different means of providing telemedicine services such as audio (ie, radio and telephone), video (ie, videoconferencing), text (ie, email), and picture (ie, still images).
- Telemedicine Services: This category defines the type of telemedicine services delivered during the study period.

Literature Evaluation

Studies were included and evaluated if and only if they presented solutions for the implementation and evaluation of telemedicine services within these scenarios. Evaluation and analysis of the included literature were conducted based on the above defined categories. The first analysis was conducted, based on the first category, to evaluate the type of circumstances (ie, maritime and offshore, extreme weather (Arctic region and Antarctica), and accident and emergency) in which the telemedicine services were delivered during the study period. A trend comparison of published literature with five years intervals was conducted to assess and compare the trends of these circumstances against the publication years. The percentages were computed based on the number of counts (n) of each type of circumstance against the publication years. For example, let us take the interval 1995-2000; the percentage was calculated based on the count of these circumstances (maritime, offshore, and remote; Arctic and extreme weather conditions; accident and emergency) addressed by the literature within this interval. The second analysis was conducted, based on the second category, to evaluate and compare the successful communication means used in these types of circumstances. The percentages were calculated based on the number of counts (n) of each subcategory of communication means used in all these types of circumstances. The third analysis was conducted, based on the third category, to evaluate the telemedicine modalities used in these types of circumstances. The percentages were computed based on the number of counts (n) of each subcategory of telemedicine modalities used in all these types of circumstances. The fourth analysis was conducted, based on the fourth category, to evaluate the type of telemedicine services delivered in the included literature in these types of circumstances. The percentages were calculated based on the number of counts (n) of each telemedicine service addressed in each type of circumstance. We noted the possibility that an article could address multiple circumstances, multiple types of communication links and modalities, and provide multiple services. Therefore, the number of characteristics reported in the Multimedia Appendix 1,Tables 1-5, and Figure 2 could exceed the number of articles reviewed.

Figure 1. Flowchart of the review process.

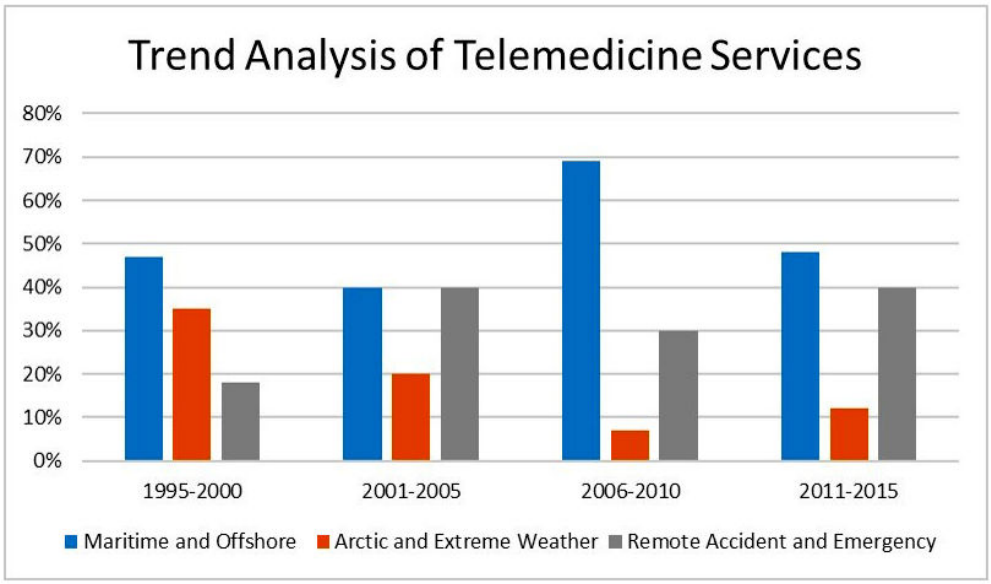


Results

Relevant Literatures

We vetted the first hit using the title, abstract, and keywords and retrieved a total of 471 papers. After removing duplicates from the group, 422 records remained. Then the authors did independent assessments of the articles, screening based on the inclusion and exclusion criteria, as illustrated in Figure 1, which eliminated another 219 papers, leaving 203 relevant papers. After a full text assessment, 36 articles were left, which were critically analyzed (as shown in Multimedia Appendix 1). The inter-rater agreement was measured using Cohen Kappa test, and disagreements were resolved through discussion. The "missing data" in Figure 1 depicts that the study either does not have a solution for the implementation, design, and evaluation of the system or is not within the context of the review.

Figure 2. Comparison of published literature in telemedicine services within the context of maritime and offshore, Arctic and extreme weather, and remote accident and emergency during the last 20 years.



Evaluation and Analysis of the Literature

The analysis and evaluation of the included articles are given in Figure 2 and Tables 1-5. The evaluation and analysis presented in this section are entirely based on the data categorizations given above.

The research effort put forward in the development of telemedicine services during the past 20 years in these different circumstances are shown in Figure 2. As shown in Figure 2, telemedicine has seen an almost steady progress in maritime and offshore services, peaking between 2006 and 2010. Moreover, telemedicine has also seen the same trend in remote accidents and emergencies as in maritime and offshore settings. However, telemedicine services in Arctic and extreme weather conditions have seen large fluctuations depicting the challenges imposed by bad weather, absence of communication networks, and absence of Arctic-enhanced telemedicine equipment in this region.

Various means of communication (networks) can be used for the implementation of telemedicine services. These include satellite, mobile (GSM, GPRS, and CDMA), ISDN, broadband, and Virtual Private Network (VPN), depending on availability and suitability. As shown in Table 1, satellite is the most used means of communication (20/70, 28%). The second most used means of communication is mobile (19/70, 27%), which includes GSM, GPRS, CDMA, and so on. Radio, VPN, and others are ranked third (13/70, 19%). ISDN is ranked fourth (11/70, 16%), followed by broadband (7/70, 10%).

Table 1. Comparison of communication links used in the published literature from 1995 to 2015.

Communication means	Usage count	Usage percentage
Satellite	20	28%
Mobile (GSM[a], GPRS[b], CDMA[c])	19	27%
Others (Radio, VPN[d], and others)	13	19%
ISDN[e]	11	16%
Broadband	7	10%

[a]GSM: Global System for Mobile Communication.

[b]GPRS: General Packet Radio Service.

[c]CDMA: code division multiple access.

[d]VPN: Virtual Private Network.

[e]ISDN: Integrated Services Digital Network.

Telemedicine services can be developed based on a different approach known as telemedicine modalities. This includes the use of video, still images, bio-signal or medical data, audio, and e-mail. As shown in Table 2, video is the most frequently used (25/93, 27%) form of delivering telemedicine. Transmission of still images is the second most used (23/93, 25%) form of delivering telemedicine. Audio (including radio and telephone) ranked third (18/93, 19%), followed by text (including email) (16/93, 17%), and others (Bio-signal transmission including ECG waveform, medical data) (11/93, 12%) ranked last.

Table 2. Comparison of published papers based on modalities of telemedicine from 1995 to 2015.

Type of modality	Usage count	Usage percentage
Video (VTC and others)	25	27%
Still pictures	23	25%
Audio (Radio, telephone, and others)	18	19%
Text (Email and others)	16	17%
Others (Bio-signal, medical data, and others)	11	12%

There are different types of onshore telemedicine services tested and evaluated in emergency and non-emergency contexts, including teleradiology, teleconsultation, teledermatology, telecardiology, telemonitoring, tele-education, tele-ENT, and others. The extent and types of such services used in the context of maritime and offshore, accident and emergency, and extreme weather (Arctic region and Antarctica) are evaluated and compared as shown in Tables 3-5.

Among the telemedicine services that are developed in maritime and offshore, teleconsultation is the most used (9/32, 28%); see Table 3. This includes both real-time (online) and offline teleconsultation that are considered and used in the studies. The second most used type of telemedicine service is Telecardiology (telesonography) (8/32, 25%). Others (7/32, 22%) takes the third rank, which consists of data sharing and decision-making, telepointing, telepresence, and others. Teledermatology (3/32, 10%) ranked fourth, followed by tele-ENT (2/32, 6%), radio-medical advice (2/32, 6%), and tele-education (1/32, 3%).

Table 3. Comparison of telemedicine services in the maritime and offshore context reported in the literature from 1995 to 2015.

Type of service	Usage count	Usage percentage
Teleconsultation	9	28%
Telecardiology (Telesonography)	8	25%
Others (Telepointing, Telepresence, Data sharing, and Decision-Making)	7	22%
Teledermatology	3	10%
Tele-ENT	2	6%
Radio medical advice	2	6%
Tele-education	1	3%

Regarding the telemedicine services in extreme weather (Arctic region and Antarctica), as shown in the Table 4, other forms of telemedicine services such as tele-interpretation, tele-ambulance, data sharing and decision-making and others are the most used forms of telemedicine (10/37, 27%). Teleconsultation is the second most used form of telemedicine (9/37, 24%) followed by teleradiology (8/37, 22%). Both teledermatology and tele-education ranked fourth most used (3/37, 8%). Tele-ENT is the fifth most used service (2/36, 6%). Telemonitoring and Telecardiology (telesonography) were ranked equally as the sixth most used telemedicine services (1/37, 3%).

Table 4. Comparison of telemedicine services in the extreme weather context reported in the literature from 1995 to 2015.

Type of service	Usage count	Usage percentage
Others (Tele-interpretation, Tele-ambulance, Data sharing and Decision-Making, and others)	10	27%
Teleconsultation	9	24%
Teleradiology	8	22%
Tele-education	3	8%
Teledermatology	3	8%
Tele-ENT	2	5%
Telemonitoring	1	3%
Telecardiology (Telesonography)	1	3%

As shown in Table 5, among the telemedicine services implemented within the context of accident and emergency, Others, which includes telepresence, tele-ophthalmology, Tele-EMS, tele-ambulance, and others are the most used means of delivering telemedicine during accident and emergency (19/60, 32%). Teleconsultation is ranked the second most used means of delivering telemedicine during accident and emergency (16/60, 27%). The third most used is telecardiology (telesonography) (7/60, 12%), followed by teleradiology (5/60, 8%), teledermatology (4/60, 7%), and telemonitoring (3/60, 5%). The least used are radio medical advice, tele-education, and Tele-ENT, each of which accounts for 3% (2/60).

Table 5. Comparison of telemedicine services in the remote accident and emergency context reported in the literature from 1995 to 2015.

Type of service	Usage count	Usage percentage
Others (Telepresence, Tele-ophthalmology, Tele-EMS, Tele-ambulance, and others)	19	32%
Teleconsultation	16	27%
Telecardiology (Telesonography)	7	12%
Teleradiology	5	8%
Teledermatology	4	7%
Telemonitoring	3	5%
Tele-ENT	2	3%
Tele-education	2	3%
Radio Medical Advice	2	3%

Discussion

Principal Findings

Telemedicine has a vital role in delivering health care services without the constraint of time and space. According to Ekeland et al [25], telemedicine has been shown to be cost effective and also to have a positive impact in various scenarios such as therapeutic effects, health care services efficiencies, and technical usability. It is an indisputable fact that telemedicine also has a transformative power on health care delivery in extreme weather, maritime/offshore, and remote emergency/accident scenarios [26]. Despite the increasing number of fishermen and other seafarers, maritime working conditions are mainly characterized by an absence of access to health care facilities. The condition is further aggravated for fishermen and seafarers who are working in the Arctic regions. Even if onshore telemedicine has been a success, its success offshore is limited. This is because of various reasons such as the absence of good communication networks, the absence of trained paramedics, bad weather conditions, and others. However, irrespective of these limitations, maritime and emergency telemedicine services have recently been successfully delivered in the Arctic, Antarctica, and other areas with extreme weather conditions. These services include teleconsultation, teleradiology, teledermatology and tele-education, telemonitoring, telecardiology (telesonography) and others (including tele-interpretation, tele-ambulance, data sharing and decision-making, and others). These services used various means of communication networks such as satellite, mobile, and radio along with different modalities such as video, still images, audio, and medical data.

Telemedicine Services in Maritime and Offshore Conditions

This section presents the status of telemedicine in maritime and offshore conditions and also identifies successful deployment of telemedicine services within these circumstances. The working conditions that exist in maritime settings are characterized by absence of access to health care facilities due to various reasons. Horneland [6] describes the limitations that hinder access to such facilities, which include the distance and time that SAR helicopters need to reach the scene, which reduces the possibility of MEDEVAC. To remedy these challenges, telemedicine is the sole choice for delivering health care services at the scene within a short response time. For instance, telemedicine services have been in use at different offshore petroleum installations and also in the Norwegian maritime fleet since 2006. Telemedicine has a great advantage for offshore personnel by providing them with access to better health care within a short response time [27]. In emergency situations, physicians onshore can thoroughly examine a patient, allowing them to accurately assess the patient's condition and develop a plan for care [28]. Moreover, telemedicine can provide advantages for companies by minimizing unnecessary medical evacuation and ship diverting for seeking medical assistance [28]. For example, Patel and Stoloff et al [29,30] conducted a cost benefit analysis of shipboard telemedicine and reported the benefit of telemedicine for a ship found at a distance of over 200 nautical miles (370 km) from shore; in such cases, the use of helicopter is found to be too costly. However, the absence of on-board trained nurses or physicians, limited equipment and medicines, availability of limited bandwidth, lack of trained paramedics, lack of complete health records, language barriers, lack of complementary medical exams, and the limitation of professional advice to only radio medical advice remains to be a challenge for the implementation of successful telemedicine services [6]. Horneland [6] clearly indicates the area that should be taken into account for the improvement of telemedicine within the context of maritime health care. Accordingly, education and training of seafarers takes priority for improvement of the services [6,31]. Additionally, preparing medical handbooks or manuals for seafarers, which are like supplements to the texts on the market, can enhance the quality of health care delivery through telemedicine. Pre-sea and periodic medical examinations are considered necessary to reduce the high risk of medical emergencies while on board [6,28]. Anscombe [28] also emphasized the need for having a proper health fitness standard that should be met before joining the task force. Recently, several research groups have performed and evaluated different clinical trials on an onshore basis and reported the success of telemedicine services [1-5]. However, this success is limited to onshore, where there is good infrastructure. By contrast, maritime telemedicine has received little attention despite the increasing number of seafarers from various nations around the world [6]. Therefore, it is deemed necessary to consider a space for adapting research results, technologies, and experiences from onshore to offshore scenarios. For example, Horneland [6] highlights the necessity of having a careful review before adapting onshore telemedicine services to offshore scenarios by justifying the case with the use of Electrocardiography (ECG) and thrombolysis. Guitton [9] gives a brief explanation about the convergence and divergence points of maritime and onshore telemedicine services by justifying the three major differences, namely structural, practical, and policy differences. Furthermore, it highlights concepts and issues for identifying these differences for better transfer of technology and research results from onshore to offshore settings. However, in spite of these differences, there are areas where direct adoption of onshore telemedicine services can be fruitful, such as patient-targeted telemedicine interventions, radio-consultation equipped with pictures and video, and video conferencing [9].

Despite the little attention maritime telemedicine has received over the past years, recently, some researchers and companies have performed studies on adopting and improving maritime telemedicine. For instance, Aujla et al [32] conducted a study on a rationalizing effort of ship to shore radio medical advice for the UK. According to Aujla et al [32], a radio medical advice is most effective when the demographic data of the population at risk are identified. Auditing the nature and frequency of medical emergencies on various types of vessels should create a basis for future recommendation regarding the minimum medical facilities needed. Furthermore, a guide for handling various conditions, that is, treatment and alternative strategies for handling the inability to evacuate patients because of bad weather should be developed. This includes provision of adequate training for the medical staff providing the services

at both sites. Likewise, Saipem's Medical Department also conducted a project to investigate and develop telecardiology so as to provide remote sites with practical support in cardiology and to extend the company's preventive approach toward cardio-vascular diseases [33]. The study has shown that telecardiology could, in fact, prevent a lot of unnecessary medical evacuations. This includes online assessment of suspected acute conditions, early detection of heart problems, and adequate filtering and priority grading of referrals for patients requiring further investigation, while reducing the load of unnecessary referrals for primary diagnosis [33]. Moreover, MERMAID is another breakthrough telemedicine project, which is a telematics-based response to the EU requirement for "long distance medical consultation" to safeguard the health and safety of maritime workers and isolated populations. The developed system is capable of delivering an integrated 24-hour multilingual worldwide emergency service to transfer medical expertise via satellite and ground-based ISDN networks [26,34-36]. The connectivity of the system is realized by combining various communication links, such as mobile satellite technologies, VSAT technologies, and ISDN protocols. The project has explored almost every category of telemedical application (audio and video conferencing, multimedia communications, flat file and image transfer with low-, medium-, and high-bandwidth data requirements), along with a full range of network choices (digital land lines, cellular or wireless, satellite, and broadband) and analysis in terms of the cost or performance trade-offs inherent to them. Moreover, it provides a variety of services, among which the notable one is electronic transmission of medical information via ISDN-based video conferencing. In addition, medical telecommunication software is considered that includes a medical record system that can guide the user through patient history and support objective examination coupled with a multimedia HELP function capability, that is, text and illustrations, based on WHO and the EU (DG V) requirements for help at sea, to guide paramedics through all the operations with a teleconsultant. Anogianakis et al [34] conducted a study that provides a means for the training and education of seafarers through the use of the MERMAID medical communications system as this is the firmest basis for the promotion of the proper practice of telemedicine at sea.

Providing access to patients' health care records can improve the health care process, thereby improving the decision-making ability of caregivers without the constraint of distance and time, whether it is at the bedside or at remote locations. One of the challenges within maritime telemedicine is the lack of complete shared health records. In this regard, Thorvik et al [12] developed and tested a telemedicine prototype known as a virtual examination room, which is an example of software for sharing medical data, enabling collaboration in different situations and based on optimal workflows between the offshore and onshore medical facilities. The concept of the virtual examination room is to give freedom and interconnectedness among the medical experts, the hospital, and the offshore nurse to simultaneously see, interpret and discuss the medical information available in the virtual examination room that has been retrieved from the connected medical devices. Likewise, Anogeianaki et al [37] also implemented a minimum medical emergency data-set (MMEDS), which enables patients to record their own health status so that information can be available for any treating physicians, irrespective of where the patients are located. The system was tested and evaluated across the Greek-Bulgarian border. Besides, Boultinghouse et al [38] reported the use of electronic medical records (EHR) during health care services delivery for oil and rig workers and implemented an electronic medical record that can keep all the health information safe, organized, and accessible over any distance, eliminating the problems and delays of a paper-based record system. This kind of shared electronic health record was provided for medics in offshore settings. Moreover, Amenta et al [31] developed an electronic medical file for each patient assisted, where the data would be updated following every radio contact with a ship or a plane in flight. Furthermore, Anogianakis et al [35] also developed software that can provide a medical record system, which can guide the user through patient history and objective examination. In addition, it provides a database that contains all the information on the vessel's stocks of medicine and medical equipment.

Assessment of user satisfaction is an important requirement for delivering quality health care services. Therefore, user satisfaction assessment on the use of maritime telemedicine has been conducted. For instance, Dehours et al [39] conducted a study on the CCMM telehealth services, the operators of French Tele-Medical Assistance Service (TMAS). During the study, 385 surveys were e-mailed, of which 165 were completed and used for analyzing user satisfaction. Overall, the result indicates that the satisfaction of on-board caregivers was high; callers were satisfied with the telephone advice, competence of physicians involved, and waiting time for services, oral prescriptions, and medical advice. The study has also given some useful recommendations for successful implementation of on-board ECG and still pictures [39]. Mair et al [40] conducted a telemedicine trial service to analyze the impact of telemedicine on reducing unnecessary evacuation. The system relies on satellite communication to provide a videoconferencing service to diagnose and treat remote oil and rig workers. The study concluded that participating onshore physicians were very satisfied on each occasion with the communications and diagnostic data and image quality, including the ultrasound screening carried out by the rig provider. The study showed that remote specialist advice via videoconferencing should reduce unnecessary and untimely patient evacuation to hospital or onshore for medical check-ups. In addition, Kevlishvili et al [41] studied the effect of teleconsultation on clinical settings. The study used videoconferencing through Skype, email, and still image services to support remote diagnosis and treatment in decision-making. Even though the trial was small, the study concluded that a telemedical solution has a great effect on simplifying remote treatment and diagnosis.

Telemedicine Services in Extreme Weather Conditions

This section presents the status of telemedicine in extreme weather condition (Arctic, Antarctica, and others) and also discusses the successful deployment of telemedicine services in these circumstances. Working in extreme cold weather has many health complications . Extreme atmospheric temperature has major consequences on the body's thermal reactions and

the risk of accidents increases when the ambient temperature falls below 0° C. Despite these health risks, currently there is a great deal of interest in the Arctic region from different companies, professionals, and merchant seafarers in connection with the discovery of huge natural resources. In order to survive in this extreme cold environment, there are services that should be put in place such as good health care services, medical examinations for fitness to work, vaccinations, first aid training for extreme cold conditions, clothing requirements, and provision of other Personal Protective Equipment (PPE) [42]. However, these regions are characterized by an absence of good health care services and weather-enhanced equipment. Due to these limitations, these regions were solely served by air ambulance operation or helicopter evacuation to get medical services from onshore specialists. However, the air ambulance operation has a lot of challenges and drawbacks including winter darkness and foggy weather. For example, Norum et al [18] analyzed air ambulance operations due to cardiovascular disease (CVD) in the Arctic region from 1999 to 2009. The study tried to analyze the challenges faced in the air ambulance operations in the Arctic region, such as long distance, rough weather conditions, and almost no alternatives for landing. According to Norum et al [18], telemedicine for remote consultation and treatment is vital for on-board vessels and rigs. However, various factors hinder the successful development of telemedicine services in the Arctic high north. This is clearly shown by Walderhaug et al [10], with a specific emphasis on the use of telemedicine in the search and rescue operation. According to Walderhaug et al [10], long distance, bad weather conditions, winter darkness, and poor communication infrastructure are some of the challenges highlighted for developing successful telemedicine services.

Despite these facts, there are groups that conducted research on alleviating these challenges. For example, the Baffin Telehealth Project is one of the Canadian High Arctic projects, which is designed to serve the remote communities of the Canadian High Arctic. The project aims at providing better health care access to the communities of the Baffin Region [43] by using different technologies from remote telemedicine systems so as to cope with isolating geography and severe environmental conditions. The system based its development on the use of a high bandwidth satellite communication to offer real-time video conferencing, digital imaging, and various medical diagnostics to support remote health stations on Baffin Island [43]. Moreover, the Mount Logan and Mount McKinley Telemedicine Projects are other examples of telemedicine projects that serve remote environments [43]. Latifi et al [44] also provide a system called the Amazon Virtual Medical Team (AVMT), which uses telemedicine services to provide health care services for swimmers in the Atlantic Ocean. The system relied on advanced technologies and a low bandwidth satellite connection to help an assembled virtual medical team to ensure telepresence 24/7 throughout the mission. Furthermore, Todnem et al [45] presented a project developed by the Statoil company for implementing telemedicine services on all Statoil operated offshore installations on the Norwegian continental shelf (NCS), succeeding in the initial pilot project from 2007-2008. The services provided included videoconferencing for meetings and educational purposes and to spread vital medical information to many locations or installations at the same time, which has been essential during epidemic situations (Swine flu, Noro virus etc). The study also demonstrated that it is possible, using the existing telemedicine equipment, to successfully remotely guide a nurse offshore in focused ultrasound examinations, with the medical doctor or expert located onshore. Similarly, one of the largest oil and gas companies, Shell, has developed remote telemedicine systems for delivering enhanced health care services to its worker [46]. The project proposes remote health care (RHC), which involves an integrated approach for delivery of health care in the Arctic operations. The system meets both the emergency and non-emergency requirements for delivering the best health care services. RHC includes different aspects such as prevention, technology, supplies or equipment, competence, and communication. The RHC serves as a virtual hospital for patients to get treated on board by allowing the physician on board the vessel to communicate real-time with onshore specialists and enabling these onshore specialists to visualize the patient using high-definition mobile cameras. The system is based on an enhanced medical technology for diagnosis and treatment, including the latest in near patient laboratory testing, digital X-Ray and “pocket” ultrasound equipment that is linked to the hospital radiology department via satellite [46]. The study assessed the outcome in Greenland, Siberia, and West Africa.

Regarding telemedicine services in the Antarctic region, various nations have performed trials for providing health care delivery for tourists and groups of researchers. For example, Grant [47] reports the experiences of using telemedicine services in the Antarctic region. According to the report, BASMU in Aberdeen have developed a tool called the Medical Assessment Questionnaire (MAQ), which proves to be effective in communicating with on-board seafarers by minimizing the treatment time, thereby allowing more accurate and error free telediagnosis and possibly reducing medical evacuations [47]. The study has also conducted various telemedicine services including successful transmission of ECG tracings through fax and e-mail for diagnosis, the potential use of thrombolysis, telemetry (even if equipment was shown to be unreliable and had poor battery life), digital x-ray equipment, ultrasound examination, internet-based education and tele-interpretation [47]. According to the report, telespirometry and more useful systems of teleconsultation are believed to be possible . Similarly, Ohno and Ohno et al [48,49] also conducted a study for delivering telemedicne solutions to be used by the Syowa Station, Japanese Antarctic Research Expedition (JARE). The developed system was intended to handle various practical cases including emergency cases. This system has shown the success of telemedicine in handling various medical operations such as surgery, orthopedics, ophthalmology, dermatology, internal medicine, urology, and dentistry [48,49]. In addition, Pillon et al [50] also reports the experiences and success of developing a telemedical solution for the principal Italian Antarctic Base at Terra Nova Bay. The system was developed to link the area with the largest Italian hospital, San Camillo in Rome. Full teleconsultation practice via videoconferencing has been developed for consultations with ophthalmic, orthopedic, and radiology specialists. Furthermore, a number of telemedicine projects have been conducted in Alaska [51], including the

Alaska Telemedicine Testbed Project (ATTP), Alaska Federal Healthcare Partnership (AFHCP), AFHCP Tele-radiology Project, the Alaska Federal Health Care Access Network (AFHCAN), AFHCAN Telemedicine Hardware and Software, and AFHCAN Connectivity & Network (WAN). Moreover, Hild [51] discusses the challenge and success factor for successful implementation of telemedicine services in Alaska. According to the report, any successful telemedicine services should be developed and designed in consideration with physical infrastructure, training structures, interoperability guidelines, and community interfaces [4]. Hence, Hild [51] also discusses the experiences of telemedicine in Alaska in accordance with these success factors.

Telemedicine Services in Remote Accident and Emergency Responses

This section presents the status of telemedicine in remote accident and emergency responses along with the search and rescue (SAR) scenarios and also discusses the successful deployment of telemedicine services (status quo) within these circumstances. Telemedicine is an optimal candidate for managing remote accidents and emergencies. However, it should be considered as a support to emergency management and not as a final solution. During an accident or emergency, the various means of communication such as real-time, store-and-forward, and data exchange can be used as either first opinion or second opinion services.

Management of accident and emergency responses are a crucial part of both onshore and offshore health care services. An accident that threatens life and health should get first aid and immediate assistance from the nearby paramedic or specialist. However, sometimes an accident could happen in a remote area such as in the Arctic region, and it takes a significant amount of time to visit a specialist. Any further delay of time in an accident means reducing the survival chances of the victims. Therefore, it is necessary to have a means of treating the victims while in a remote area, at least to prolong the chance of survival. The telemedicine system designed for real-time emergencies has paved the way for a new perspective in remote medical diagnosis [52]. Ensuring safety in the Arctic waters is very challenging because of the remoteness of the region and the lack of contingency planning infrastructure [4]. According to Berg et al [53], so as to reduce the rate of accidents in the Arctic waters, it is necessary to improve regional collaboration, develop additional professional requirements for seafarers, and provide training in order to offer knowledge transfer from seniors working in the region. For example, Buschmann et al [54] described the need for the medical education concept, “SAR-First Responder Sea,” to help paramedics in providing treatment and diagnosis during search and rescue operations. Miller et al [55] also conducted a retrospective review to analyze the potential of emergency nurse practitioners (ENPs) for delivering telemedicine advice for minor injuries. The result is in agreement with Buschmann et al [54], which supports the conclusion that the assessment of all minor injuries through a telemedicine network by medical staff is unnecessary and that an ENP-led service offers a realistic and attractive alternative. Similarly, Boniface et al [56] conducted a study to assess the capability of ultrasound-naive paramedics to obtain interpretable Focused Assessment with Sonography for Trauma (FAST) pictures under the remote guidance of Emergency Physicians (EPs). The result has shown that paramedics with no prior ultrasound experience could obtain FAST images under remote guidance from experienced EPs in less than 5 minutes. This result has a potential advantage for managing remote accidents by treating patients through data transmission [56]. In addition, Bergrath et al [57] investigated the feasibility and effect of pre-hospital teleconsultation to transfer the concept into the emergency medical services. The study was conducted in a real clinical setting by comparing telemedically assisted pre-hospital care (telemedicine group) with the local regular EMS care (control group). The study concluded that teleconsultation is feasible but technical performance and reliability have to be improved. Moreover, the result has shown the future potential of pre-hospital tele-stroke consultation to improve emergency care, especially when no highly trained personnel are on-scene [57]. Furthermore, Brebner et al [58] evaluated a pilot telemedicine network for accident and emergency work. The study assessed the treatment and diagnosis of emergency cases through videoconferencing for a period of 15 months. The study demonstrated that accident and emergency teleconsultations can be technically reliable and effective in reducing the number of patient transfers and in a manner acceptable to the referring clinicians. Similarly, Bowman et al [59] conducted a controlled trial to assess the accuracy of telemedicine in diagnosing and managing eye problems presented to accident and emergency. The study has shown that telemedicine, utilizing video slit lamp images, to be an effective, safe, and accurate method of diagnosing and managing these patients.

Castellano et al [52] conducted a study for handling emergency cases in a pre-hospital environment. The study designed a real-time emergency telemedicine system for remote medical diagnosis (an ambulance) using a hybrid network that enabled secure long-distance communication from an ambulance. Moreover, the study demonstrated a specific scenario by performing an ambulance-based hematological test with regard to an international normalized ratio (INR) using wireless transmission, accurately and in real-time, to the referral hospital. The study reported no significant differences between the ambulance-based and the laboratory-based tests [52].

Kang et al [60] conducted a preliminary study that evaluated the use of a code division multiple access (CDMA)-based emergency telemedicine system to be used by emergency rescuers providing first-aid treatment to patients. The evaluated prototype consisted of equipment for measuring non-invasive arterial blood pressure (NIBP), arterial oxygen saturation (SpO2), six-channel electrocardiogram (ECG), blood glucose concentration, and body temperature. The recorded patient data were transmitted to the doctor’s computer through CDMA and TCP/IP networks using an embedded personal digital assistant (PDA) phone. The result indicates that the systems provided reliable values. Moreover, the feasibility of the prototype was evaluated with 15 real emergency patients on Jeju Island over a two-month period. The measured data were successfully transmitted without significant CDMA connection loss or transmission errors.

Uldal et al [61] developed a mobile telemedicine unit (MTU) for emergency and screening purposes, which included various facilities such as endoscopy, electrocardiography, and digital photography. The mobile telemedicine unit included a modem, a portable PC, an endoscope, an electrocardiogram (ECG) facility, a digital camera, a printer, and an uninterruptible power supply (UPS). In this system, data transmission was implemented using an ordinary telephone line.

Kyriacou et al [62] developed a portable medical device that supported emergency telemedicine by allowing telediagnosis, long distance support, and teleconsultation of mobile health care providers by expert physicians. The system combined both real-time and store and forward facilities by using a telemedicine unit at the patient or emergency site and the expert's medical consulting at the base unit. This integrated system had a capability to be used in various emergency cases such as being treated in an ambulance vehicle, in a Rural Health Centre and on a navigating ship. The developed system was a "multi-purpose" telemedicine system consisting of two major parts: a telemedicine unit (located near the patient) and a base unit (located at a Central Hospital). The Telemedicine unit mainly consisted of hardware and software components such as a bio-signal acquisition module, image capturing module, main module, and communication module. The system relied on a GSM or GPRS modem and a POTS modem or satellite modem for communication purposes. The system was clinically evaluated and installed and used in two different countries: Greece (ambulance vehicles, rural health centers, ships) and Cyprus (ambulance vehicles, rural health centers) [62]. The study also performed a comparison regarding the performance of communication media such as GSM and satellite, which indicates that satellite communication performs well when handling a large file size [62].

Conclusions

Despite the increasing number of professionals, fishermen, and merchant seafarers, maritime working conditions are characterized by an absence of access to health care facilities. This condition is further aggravated for seafarers who are working in the Arctic regions. Even if telemedicine has seen success onshore, there is limited success to it offshore. This is due to the absence of a good communication network, lack of trained personnel (paramedics), lack of maritime and offshore enhanced telemedicine equipment, bad weather conditions, and the distance and time required for SAR helicopters to reach a site, which reduces the possibility of MEDEVAC. Technology adoption from onshore to offshore settings might seem like a quick remedy for the case, but this remains a challenge for various reasons. The major challenge lies in the convergent and divergent nature of maritime and onshore telemedicine with respect to structure, practice, and policy. Therefore, it is necessary to identify these divergent and convergent points and to carefully review them before transferring technology and research experiences to offshore scenarios.

Despite these limitations, recently, a number of successes have been achieved in delivering telemedicine services in maritime settings, remote emergency responses, and Arctic and other extreme weather scenarios. These services include teleconsultation, teleradiology, telecardiology, tele-ENT, teledermatology, and tele-education, to mention a few. Most of these studies demonstrate the use of various means of communication including satellite, mobile, radio, and others. Moreover, all these studies have shown the use of various telemedicine modalities including video, still images, audio, and medical data. However, the use of telemedicine in relation to the search and rescue (SAR) services is not yet fully exploited. During this review, we did not see a paper that implemented telemedicine services for search and rescue (SAR) scenarios. Therefore, we foresee that once implemented and evaluated, these telemedicine services will serve as an underlying model for the successful deployment of the future telemedicine-assisted search and rescue (SAR) services. Even though the provision of telemedicine within the extreme weather (Arctic) and offshore scenarios calls for having good communication infrastructure and more Arctic-enhanced equipment, successful telemedicine services cannot be met with only these technologies. It needs an organization that is committed, motivated, and willing to invest in a project, while also being capable of mobilizing the human performance factors toward delivering the services.

Practice Points

The following points are worth considering for a meaningful telemedicine solution in the context of maritime and extreme weather conditions as in the Arctic region:

First, maritime and onshore telemedicine can be convergent and divergent with respect to structural, practical, and policy differences. Therefore, it is necessary to identify these differences and carefully review them before transferring technology and research experiences to offshore scenarios.

Second, sometimes evacuation might become difficult in the Arctic region; therefore, it is necessary to consider telemedicine as the actual health care delivery services, rather than simply considering it as means of information exchange.

Third, it is necessary to do more Arctic-enhanced telemedicine research and also to assess and analyze the telemedicine solutions deployed in other regions such as Antarctica.

Fourth, it is necessary to have a systematic analysis of previous accidents in the Arctic region in order to provide a knowledge base with respect to emergency preparedness and response, focusing on the various phases and types of accidents.

Fifth, it is important to analyze the status quo of search and rescue operation in the Arctic region, so as to identify the capability gaps and to take the necessary measures.

Sixth, as applying the current open water Escape, Evacuation, and Rescue technology might have an unacceptably high failure rate, it is necessary to deploy more Arctic-enhanced EER technology.

Seventh, in most of the literature, we have observed the lack of a common international standard or protocol for information sharing, that is, DICOM (Digital Imaging and Communication in Medicine). However, we have noticed some articles [50,62] that were based on the development of "Vital" and "DICOM"

standards. Therefore, it is necessary to adopt such kinds of standard for an interoperable information communication.

Finally, it is necessary to consider the capability approach in the broad concept of integrated operation for offshore telemedicine services.

Acknowledgments

We would like to thank our colleagues from the "Search and Rescue in the high North" (SARINOR) project, work package 4/5, for their invaluable inputs and support throughout the study. For further information about SARINOR reports, please visit http://www.sarinor.no. We would also like to thank UiT (The Arctic University of Norway) for funding the publication processing fee.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Literature included in the study, sorted from the latest to the oldest.

[PDF File (Adobe PDF File), 86KB - medinform_v5i2e16_app1.pdf]

References

1. Hartvigsen G, Pedersen S. Lessons learned from 25 years with telemedicine in Tromsø, Norway. Tromsø, Norway: Norwegian Centre for Integrated Care and Telemedicine, University Hospital of North Norway; 2015.
2. Hartvigsen G, Johansen MA, Hasvold P, Bellika JG, Arsand E, Arild E, et al. Challenges in telemedicine and eHealth: lessons learned from 20 years with telemedicine in Tromsø. Stud Health Technol Inform 2007;129(Pt 1):82-86. [Medline: 17911683]
3. Nesbitt TS, Dharmar M, Katz-Bell J, Hartvigsen G, Marcin JP. Telehealth at UC Davis--a 20-year experience. Telemed J E Health 2013 May;19(5):357-362. [doi: 10.1089/tmj.2012.0284] [Medline: 23343257]
4. Hild M. Arctic Council. 2000. Arctic telemedicine project final report URL: https://oaarchive.arctic-council.org/handle/11374/27 [accessed 2017-05-26] [WebCite Cache ID 6qkdMfoKR]
5. Pedersen S, Gammon D, Sund T. Telemedicine: a means to improve health care services on remote oil and gas installations. 1994 Presented at: SPE Health, Safety and Environment in Oil and Gas Exploration and Production Conference; January 25-27,1994; Jakarta, Indonesia URL: https://www.onepetro.org/download/conference-paper/SPE-27201-MS?id=conference-paper%2FSPE-27201-MS [doi: 10.2118/27201-ms]
6. Horneland AM. Maritime telemedicine - where to go and what to do. Int Marit Health 2009;60(1-2):36-39. [Medline: 20205126]
7. Guitton MJ. Online maritime health information: an overview of the situation. Int Marit Health 2015;66(3):139-144 [FREE Full text] [doi: 10.5603/IMH.2015.0028] [Medline: 26394313]
8. Duffy B. Dental problems in the offshore oil and gas industry: a review. Occup Med (Lond) 1996 Feb;46(1):79-83. [Medline: 8672801]
9. Guitton MJ. Telemedicine at sea and onshore: divergences and convergences. Int Marit Health 2015;66(1):18-21 [FREE Full text] [doi: 10.5603/IMH.2015.0005] [Medline: 25792161]
10. Walderhaug S, Granja C, Horsch A, Hartvigsen G. Telemedicine Services in Arctic Environments-Challenges for Successful Implementation. In: Linköping Electronic Conference Proceedings. Linköping, Sweden: Linköping University Electronic Press; 2015 Presented at: The 13th Scandinavian Conference on Health Informatics; June 15-17, 2015; Tromsø, Norway p. 98-101.
11. Basharat S. Accidents and emergency response in the Arctic sea. In: OTC Arctic Technology Conference. Offshore Technology Conference.: Offshore Technology Conference; 2014 Presented at: OTC Arctic Technology Conference; February 10-12, 2014; Houston, Texas URL: https://www.onepetro.org/download/conference-paper/OTC-24609-MS?id=conference-paper%2FOTC-24609-MS [doi: 10.4043/24609-MS]
12. Thorvik K, Nystad A, Fernandes A, Reegard K, Simensen J, Rindahl G, et al. The future of telemedicine in O&G. : Society of Petroleum Engineers; 2014 Presented at: SPE Intelligent Energy Conference & Exhibition; April 1-3, 2014; Utrecht, The Netherlands URL: https://www.onepetro.org/download/conference-paper/SPE-167841-MS?id=conference-paper%2FSPE-167841-MS [doi: 10.2118/167841-ms]
13. Rottem SV. The Arctic council and the search and rescue agreement: the case of Norway. Polar Record 2013 May 13;50(03):284-292. [doi: 10.1017/s0032247413000363]
14. Lloyd C, Perry J. A review of international environmental and social guidelines for offshore Arctic oil and gas activities, current requirements and future trends. : Society of Petroleum Engineers; 2013 Presented at: SPE Arctic and Extreme

Environments Technical Conference and Exhibition; 2013; Moscow, Russia URL: https://www.onepetro.org/download/conference-paper/SPE-166977-MS?id=conference-paper%2FSPE-166977-MS [doi: 10.2118/166977-ms]
15. Ims B. Bibsys. Trondheim: Norwegian University of Science and Technology; 2013. Emergency preparedness in Arctic oil and gas exploration URL: https://brage.bibsys.no/xmlui/handle/11250/238425?show=full [accessed 2017-06-05] [WebCite Cache ID 6qzNAzGUy]
16. Barnes RJ. The challenges of russian Arctic projects. : Society of Petroleum Engineers; 2011 Presented at: SPE Arctic and Extreme Environments Conference and Exhibition; October 18-20, 2011; Moscow, Russia URL: https://www.onepetro.org/download/conference-paper/SPE-149574-MS?id=conference-paper%2FSPE-149574-MS [doi: 10.2118/149574-ms]
17. Haagensen R, Sjøborg K, Rossing A, Ingilae H, Markengbakken L, Steen P. Long-range rescue helicopter missions in the Arctic. Prehosp Disaster Med 2004;19(2):158-163. [Medline: 15508199]
18. Norum J. Cardiovascular disease (CVD) in the Norwegian Arctic. Air ambulance operations 1999-2009 and future challenges in the region. Int Marit Health 2010;62(3):117-122. [Medline: 21154297]
19. Marsden A, Totten M, Spring W. Feasibility of escape, evacuation and rescue for facilities in Arctic shear zone environments. : Offshore Technology Conference; 2011 Presented at: OTC Arctic Technology Conference; February 7-9, 2011; Houston, TX p. 2013 URL: https://www.onepetro.org/download/conference-paper/OTC-22055-MS?id=conference-paper%2FOTC-22055-MS [doi: 10.4043/22055-ms]
20. Steinicke S, Albrecht S. SWP Berlin. 2012. Search and rescue in the Arctic URL: https://www.swp-berlin.org/fileadmin/contents/products/arbeitspapiere/WP_FG2_2012_Steinicke_Albrecht.pdf [accessed 2017-06-05] [WebCite Cache ID 6qzNmhr7x]
21. Benger J. A review of telemedicine in accident and emergency: the story so far. J Accid Emerg Med 2000 May;17(3):157-164 [FREE Full text] [Medline: 10819375]
22. Keane MG. A review of the role of telemedicine in the accident and emergency department. J Telemed Telecare 2009;15(3):132-134. [doi: 10.1258/jtt.2009.003008] [Medline: 19364895]
23. Amadi-Obi A, Gilligan P, Owens N, O'Donnell C. Telemedicine in pre-hospital care: a review of telemedicine applications in the pre-hospital environment. Int J Emerg Med 2014;7:29 [FREE Full text] [doi: 10.1186/s12245-014-0029-0] [Medline: 25635190]
24. Ward MM, Jaana M, Natafgi N. Systematic review of telemedicine applications in emergency rooms. Int J Med Inform 2015 Sep;84(9):601-616. [doi: 10.1016/j.ijmedinf.2015.05.009] [Medline: 26072326]
25. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
26. Anogianakis G, Maglavera S, Pomportsis A. Relief for maritime medical emergencies through telematics. IEEE Trans Inf Technol Biomed 1998 Dec;2(4):254-260. [Medline: 10719535]
27. Fernandes A, Reegård K, Drøivoldsmo A, Simensen J, Rindahl G. Development of telemedicine in oil & gas through the capabilities approach. 2014 Presented at: The 5th International Conference on Applied Human Factors and Ergonomics (AHFE); July 19-23, 2014; Kraków, Poland p. 19-23 URL: https://www.researchgate.net/publication/271485380_Development_of_telemedicine_in_oil_gas_through_the_capabilities_approach
28. Anscombe DL. Healthcare delivery for oil rig workers: telemedicine plays a vital role. Telemed J E Health 2010;16(6):659-663. [doi: 10.1089/tmj.2010.9957] [Medline: 20690836]
29. Patel T. A cost-benefit analysis of the effect of shipboard telemedicine in a selected oceanic region. J Telemed Telecare 2000;6(Suppl 1):S165-S167. [Medline: 10794008]
30. Stoloff PH, Garcia FE, Thomason JE, Shia DS. A cost-effectiveness analysis of shipboard telemedicine. Telemed J 1998;4(4):293-304. [doi: 10.1089/tmj.1.1998.4.293] [Medline: 10220469]
31. Amenta F, Capone L, Sibilio F. CIRM. 2013. Telemedical Assistance of Patients on Board Ships Activity of Centro Internazionale Radio Medico (CIRM), the Italian Telemedical Maritime Assistance Service (TMAS) URL: http://www.cirm.it/documenti/cirm_th3-1.pdf [accessed 2017-06-05] [WebCite Cache ID 6qzOLTH3p]
32. Aujla K, Nag R, Ferguson J, Howell M, Cahill C. Rationalizing radio medical advice for maritime emergencies. J Telemed Telecare 2003;9(Suppl 1):S12-S14. [doi: 10.1258/135763303322196178] [Medline: 12952706]
33. Panait D, Mika F. Tele-cardiology in remote O&G premises. : Society of Petroleum Engineers; 2013 Presented at: European HSE Conference and Exhibition; April 16-18, 2013; London, United Kingdom URL: https://www.onepetro.org/download/conference-paper/SPE-164984-MS?id=conference-paper%2FSPE-164984-MS [doi: 10.2118/164984-ms]
34. Anogianakis G, Maglavera S. Utilising multimedia for training merchant mariners as paramedics. Stud Health Technol Inform 2000;72:66-72. [Medline: 11010337]
35. Anogianakis G, Maglavera S, Pomportsis A, Bountzioukas S, Beltrame F, Orsi G. Medical emergency aid through telematics: design, implementation guidelines and analysis of user requirements for the MERMAID project. Int J Med Inform 1998;52(1-3):93-103. [Medline: 9848406]
36. Anogianakis G, Maglavera S. Medical emergency aid through telematics (MERMAID). Stud Health Technol Inform 1996;29:255-264. [Medline: 10163758]
37. Anogeianaki A, Papaliagkas V, Guibas G, Anogianakis G. Telemedicine services across the Greek-Bulgarian border. J Telemed Telecare 2007 Dec 01;13(8):1-3. [doi: 10.1258/135763307783247248]

38. Boultinghouse O, Fitts JT. Drillingcontractor. 2009. Telemedicine technologies enhance offshore healthcare, reduce illness-related departures URL: http://www.drillingcontractor.org/telemedicine-technologies-enhance-offshore-healthcare-reduce-illness-related-departures-1853 [accessed 2017-05-26] [WebCite Cache ID 6qkkKUSKg]
39. Dehours E, Vallé B, Bounes V, Girardi C, Tabarly J, Concina F, et al. User satisfaction with maritime telemedicine. J Telemed Telecare 2012 Jun;18(4):189-192. [doi: 10.1258/jtt.2012.110910] [Medline: 22604271]
40. Mair F, Fraser S, Ferguson J, Webster K. Telemedicine via satellite to support offshore oil platforms. J Telemed Telecare 2008;14(3):129-131. [doi: 10.1258/jtt.2008.003008] [Medline: 18430278]
41. Kevlishvili G, Mika F, Sanctis SD. Trial teleconsultation sessions on Castoro 12 in North Caspian sea. : Society of Petroleum Engineers; 2011 Presented at: SPE European Health, Safety and Environmental Conference in Oil and Gas Exploration and Production; February 22-24, 2011; Vienna, Austria URL: https://www.onepetro.org/download/conference-paper/SPE-140501-MS?id=conference-paper%2FSPE-140501-MS [doi: 10.2118/140501-ms]
42. Barbey A, Covil M, Dahl-Hansen E, Dawson D, de JG, Dugelay F, et al. Health aspects of work in extreme cold within the E & P industry. : Society of Petroleum Engineers; 1998 Presented at: SPE International Conference on Health, Safety, and Environment in Oil and Gas Exploration and Production; June 7-10, 1998; Caracas, Venezuela URL: https://www.onepetro.org/download/conference-paper/SPE-46685-MS?id=conference-paper%2FSPE-46685-MS [doi: 10.2118/46685-ms]
43. Otto CA. Telemedicine in the Canadian high Arctic and other remote environments. In: Engineering in Medicine and Biology 21st Annual Conference and the 1999 Annual Fall Meetring of the Biomedical Engineering Society BMES/EMBS Conference.: IEEE; 1999 Presented at: Proceedings of the First Joint BMES/EMBS Conference. 1999 IEEE Engineering in Medicine and Biology 21st Annual Conference and the 1999 Annual Fall Meeting of the Biomedical Engineering Society (Cat. N; October 1999; Atlanta, GA URL: http://ieeexplore.ieee.org/document/803864/ [doi: 10.1109/IEMBS.1999.803864]
44. Latifi R, Stanonik M, Merrell RC, Weinstein RS. Telemedicine in extreme conditions: supporting the Martin Strel Amazon Swim Expedition. Telemed J E Health 2009 Jan;15(1):93-100. [doi: 10.1089/tmj.2008.0057] [Medline: 19199853]
45. Todnem K, Evensen A, Oveland N. The Implementation of Telemedicine As An Integrated Part of The Health Service on The Statoil Operated Installations on The Norwegian Continental Shelf (NCS). : Society of Petroleum Engineers; 2012 Presented at: The implementation of telemedicine as an inteInternational Conference on Health, Safety and Environment in Oil and Gas Exploration and Production; September 11-13, 2012; Perth, Australia URL: https://www.onepetro.org/download/conference-paper/SPE-157562-MS?id=conference-paper%2FSPE-157562-MS [doi: 10.2118/157562-ms]
46. Berg J, Toner S, Stilz R, Klein S, Williams H, Pearson J, et al. Remote health care: a game changer for the Arctic. : Offshore Technology Conference; 2015 Presented at: OTC Arctic Technology Conference; March 23-25, 2015; Copenhagen, Denmark URL: https://www.onepetro.org/download/conference-paper/OTC-25549-MS?id=conference-paper%2FOTC-25549-MS [doi: 10.4043/25549-ms]
47. Grant IC. Telemedicine in the British Antarctic survey. Int J Circumpolar Health 2004 Dec;63(4):356-364. [Medline: 15709312]
48. Ohno G. Practical results of telemedicine system between Antarctic station and Japan. In: Graschew G, Rakowsky S, editors. Telemedicine Techniques and Applications. Croatia - European Union: INTECH Open Access Publisher; 2011.
49. Ohno G, Watanabe K, Okada Y, Higuchi K. Practical experience of telehealth between an Antarctic station and Japan. J Telemed Telecare 2012 Dec;18(8):473-475. [doi: 10.1258/jtt.2012.GTH111] [Medline: 23209273]
50. Pillon S, Todini AR. eHealth in Antarctica: a model ready to be transferred to every-day life. Int J Circumpolar Health 2004 Dec;63(4):436-442. [Medline: 15709319]
51. Hild C. Arctic telehealth: north to the future. Int J Circumpolar Health 2004;63(Suppl 2):63-70. [Medline: 15736624]
52. Castellano NN, Gazquez JA, García Salvador RM, Gracia-Escudero A, Fernandez-Ros M, Manzano-Agugliaro F. Design of a real-time emergency telemedicine system for remote medical diagnosis. Biosyst Eng 2015 Oct;138:23-32. [doi: 10.1016/j.biosystemseng.2015.03.017]
53. Berg TE, Kvamstad B, Kjersem F. Safety at sea – a review of Norwegian activities. TransNav 2011;5(2):195-201 [FREE Full text]
54. Buschmann C, Niebuhr N, Schulz T, Fox UH. “SAR-First-Responder Sea” - backgrounds to a medical education concept in German SAR service. Int Marit Health 2009;60(1-2):43-47. [Medline: 20205128]
55. Miller DR, Alam K, Fraser S, Ferguson J. The delivery of a minor injuries telemedicine service by emergency nurse practitioners. J Telemed Telecare 2008;14(3):143-144. [doi: 10.1258/jtt.2008.003013] [Medline: 18430283]
56. Boniface KS, Shokoohi H, Smith ER, Scantlebury K. Tele-ultrasound and paramedics: real-time remote physician guidance of the focused assessment with sonography for trauma examination. Am J Emerg Med 2011 Jun;29(5):477-481. [doi: 10.1016/j.ajem.2009.12.001] [Medline: 20825815]
57. Bergrath S, Reich A, Rossaint R, Rörtgen D, Gerber J, Fischermann H, et al. Feasibility of prehospital teleconsultation in acute stroke--a pilot study in clinical routine. PLoS One 2012;7(5):e36796 [FREE Full text] [doi: 10.1371/journal.pone.0036796] [Medline: 22629331]
58. Brebner EM, Brebner JA, Ruddick-Bracken H, Wootton R, Ferguson J. Evaluation of a pilot telemedicine network for accident and emergency work. J Telemed Telecare 2002;8(Suppl 2):5-6. [doi: 10.1258/135763302320301803] [Medline: 12217114]

59. Bowman RJ, Kennedy C, Kirwan JF, Sze P, Murdoch IE. Reliability of telemedicine for diagnosing and managing eye problems in accident and emergency departments. Eye (Lond) 2003 Aug;17(6):743-746. [doi: 10.1038/sj.eye.6700489] [Medline: 12928688]
60. Kang J, Chun H, Shin IH, Shin SD, Suh GJ, Kim HC. Preliminary evaluation of the use of a CDMA-based emergency telemedicine system. J Telemed Telecare 2006;12(8):422-427. [doi: 10.1258/135763306779378744] [Medline: 17227609]
61. Uldal SB, Amerkhanov J, Manankova BS, Mokeev A, Norum J. A mobile telemedicine unit for emergency and screening purposes: experience from north-west Russia. J Telemed Telecare 2004;10(1):11-15. [doi: 10.1258/135763304322764121] [Medline: 15006209]
62. Kyriacou E, Pavlopoulos S, Koutsouris D. An emergency telemedicine system based on wireless communication technology: a case study. In: Istepanian RS, Laxminarayan S, Pattichis C, editors. M-health: emerging mobile health systems. New York, NY: Springer; 2006:401-416.

Abbreviations

CDMA: code division multiple access
DICOM: Digital Imaging and Communication in Medicine
EER: Escape, Evacuation, and Rescue
ENPs: emergency nurse practitioners
GSM: Global System for Mobile Communication
GPRS: General Packet Radio Service
ISDN: Integrated Services Digital Network
MEDEVAC: medical evacuation
RHC: remote health care
SAR: search and rescue
VPN: Virtual Private Network
VSAT: very small aperture terminal

Edited by G Eysenbach; submitted 06.07.16; peer-reviewed by M Guitton, D Seale; comments to author 28.07.16; revised version received 07.10.16; accepted 04.04.17; published 28.06.17.

Please cite as:
Woldaregay AZ, Walderhaug S, Hartvigsen G
Telemedicine Services for the Arctic: A Systematic Review
JMIR Med Inform 2017;5(2):e16
URL: http://medinform.jmir.org/2017/2/e16/
doi:10.2196/medinform.6323
PMID:28659257

Original Paper

Virtual Reality as an Adjunct Home Therapy in Chronic Pain Management: An Exploratory Study

Bernie Garrett[1*], RN, PhD; Tarnia Taverner[1*], RN, PhD; Paul McDade[2], BSc

[1]School of Nursing, University of British Columbia, Vancouver, BC, Canada

[2]Faculty of Science, University of British Columbia, Vancouver, BC, Canada

[*]these authors contributed equally

Corresponding Author:
Bernie Garrett, RN, PhD
School of Nursing
University of British Columbia
T 201 2211 Wesbrook Mall
Vancouver, BC, V6T 2B5
Canada
Phone: 1 604 753 8942
Fax: 1 604 822 7466
Email: bernie.garrett@nursing.ubc.ca

Abstract

Background: Virtual reality (VR) therapy has been successfully used as an adjunct therapy for the management of acute pain in adults and children, and evidence of potential efficacy in other health applications is growing. However, minimal research exists on the value of VR as an intervention for chronic pain.

Objective: This case series examined the value of VR to be used as an adjunctive therapy for chronic pain patients in their own homes.

Methods: An exploratory approach using a case series and personal interviews was used. Ten chronic pain patients received VR therapy for 30 min on alternate days for 1 month. Pre- and postexposure (immediately afterwards, 3 h, and at 24 h) pain assessment was recorded using the Numerical Rating Scale (NRS), and weekly using the Brief Pain Inventory (BPI) and Self-completed Leeds Assessment of Neuropathic Symptoms and Signs pain scale (S-LANSS). Terminal semistructured personal interviews with the patients were also undertaken.

Results: Of the 8 patients who completed the study, 5 of them reported that pain was reduced during the VR experience but no overall treatment difference in pain scores postexposure was observed. VR was not associated with any serious adverse events, although 60% of patients reported some cybersickness during some of the experiences.

Conclusions: Of note is that the majority of these study participants reported a reduction in pain while using the VR but with highly individualized responses. One patient also reported some short-term improved mobility following VR use. Some evidence was found for the short-term efficacy of VR in chronic pain but no evidence for persistent benefits.

(JMIR Med Inform 2017;5(2):e11) doi:10.2196/medinform.7271

KEYWORDS

pain management; chronic illness; therapeutics; medical informatics

Introduction

Research on virtual reality (VR) dates back to the early 1980s [1], but the potential for mainstream use has only recently been realized. Some successes have been reported in the use of VR in the treatment of acute pain as an adjunctive method for pain control [2-7]. Clinical studies exploring its use for chronic pain remain minimal [8-10]. Results have been hopeful, but some used artificially induced pain in healthy adults rather than actual chronic pain patients; thus, findings may not be clinically comparable [8]. Nevertheless, the value of VR in chronic pain management remains an area of potentially high impact research. In Canada, chronic pain is a significant health issue with 18.9% of adult Canadians suffering from chronic persistent pain [11]. Chronic pain persists as a complex phenomenon affecting millions of Canadians every day [11-15]. Chronic pain patients

also often find their pain experiences persist despite medical interventions, and those affected frequently suffer from additional decreases in psychosocial health and activity restriction [16-17].

Cognitive factors are well-known to affect perceptions of pain [18]. Currently, VR environments are hypothesized to reduce pain via cognitive attentional and distractive mechanisms, although the exact mechanisms remain unclear [19-25]. The use of VR might act directly and indirectly on pain perception in a number of ways by altering signaling pathways involving attention, emotion, concentration, memory, touch, and the auditory and visual senses. VR interventions appear to reduce pain sensitivities by altering the sense of personal presence to that of being in new virtual environment, changing the sensory, affective, and cognitive features of the experience and altering the subjective perception of pain [26,27]. Therefore, the potential value of VR to help mediate chronic pain is an important area for exploration. The primary aims for this exploratory study were to identify any changes from baseline pain scores and in reported pain experiences using VR, and establish whether VR can be practically and safely used at home. Secondary aims included identifying any weekly pain score changes, any adverse effects [28], effects on function, and any preferences in type of VR experience. In addition, the study was undertaken to evaluate feasibility and establish practical methods to research VR interventions for chronic pain.

Methods

Design

A mixed-methods pilot case-series approach was used. The quantitative a priori hypotheses tested were that for patients with established chronic pain treated at home: (1) exposure to VR for 30 min 3 times a week would decrease pain scores from their preexposure baseline, and (2) exposure to VR sessions 3 times a week would lead to decreased weekly pain scores over a month. The qualitative aspects of the study examined patient's perceptions of their pain while using VR, if they observed any practical application of safety issues in using VR or experienced any adverse effects. Also, their VR experience preferences, and if they noted any functional or quality-of-life improvements during the study were examined.

Sample Selection and Recruitment

Prior to recruitment, a review of the proposal was undertaken by the UBC Clinical Research Ethics Board and approval granted. A purposeful nonprobability convenience sample of 10 adult patients with a diagnosed chronic pain condition for at least six months were recruited by Web-based invitations from 2 sources; PainBC and the People in Pain Network. Both are charitable support organizations, based in British Columbia. Interested patients were sent further information and a telephone interview by a member of the research team was conducted to answer the patients' questions and screen them against specific inclusion and exclusion criteria (see Textboxes 1 and 2). Suitable patients were then sent an informed consent form to be signed and returned on installation of the VR equipment (with a duplicate copy provided for them to retain).

Textbox 1. Inclusion criteria for the study.

- ≥18 years
- Have had a chronic pain diagnosis for 6 months or longer
- Score a maximum of ≥4 on the NRS pain scale daily
- Have desk space at home for the VR headset and accompanying computer system
- Able to understand the English language, and read and write English
- Able to wear a VR HMD (head-mounted display) and move head in cervical rotation, extension, and flexion
- sufficient fine motor control to operate a joystick/game controller

Textbox 2. Exclusion criteria for the study.

- Individuals who have cognitive impairment or inability to control a basic computer VR interface, or complete questionnaires
- Susceptibility to motion sickness or cyber-sickness (LaViole 2001)
- Susceptibility to claustrophobia
- History of susceptibility to seizures

Intervention

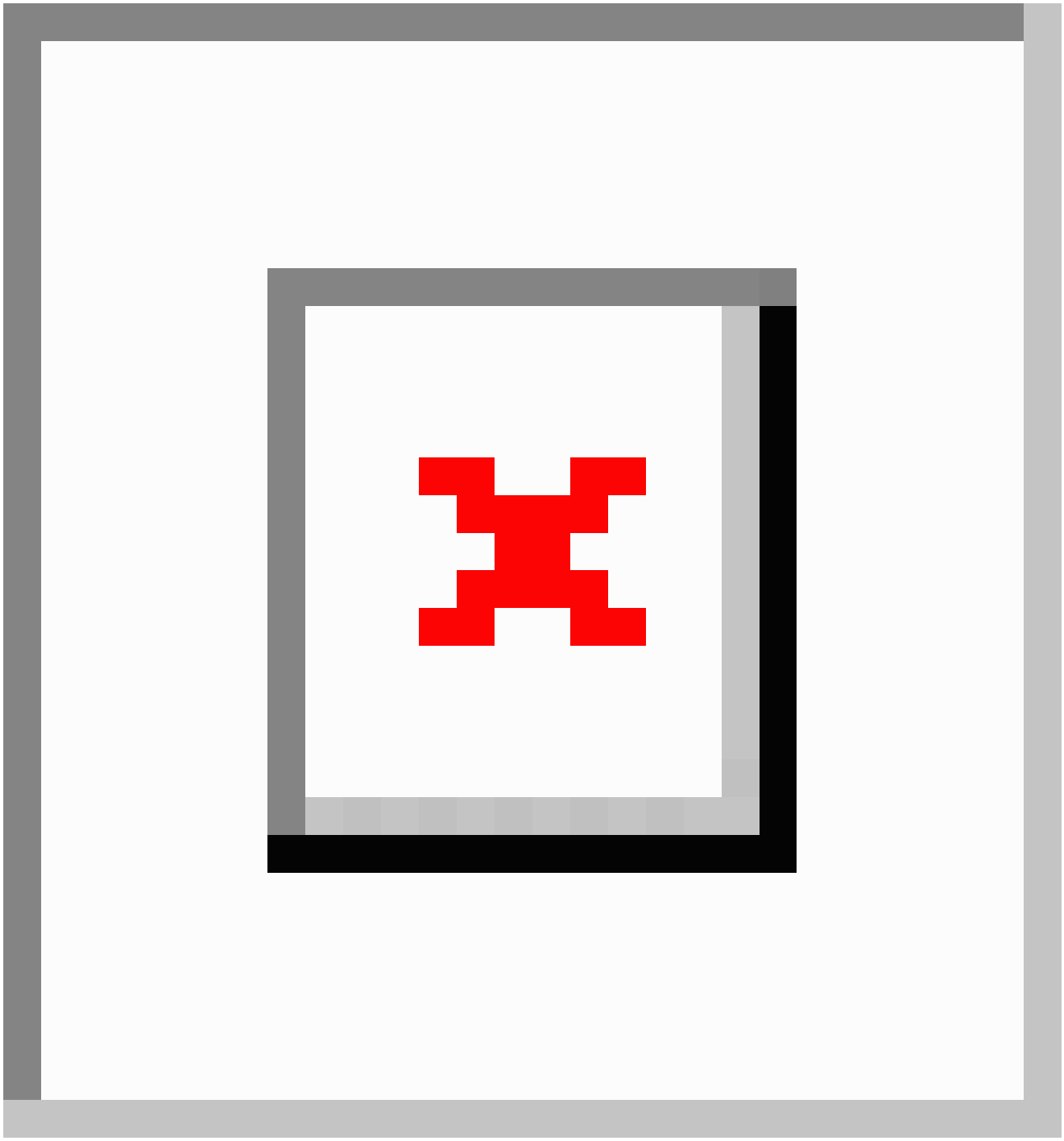
A home-based VR intervention was selected for the study. First, for practicality, as many of these patients also had mobility concerns and it was not economical to offer transportation to a hospital or lab and back again several times a week. Second, commercial VR systems would need to be suitable and easy to use at home, if they proved efficacious in the treatment of chronic pain. The VR hardware used was identical for each patient, and consisted of a high-end personal computer running the VR applications with an Oculus Rift DK2 110^0 field of view (FOV) stereoscopic head-mounted display (HMD) with a resolution of 960 × 1080 pixels per eye.

As there were no VR experiences validated in the context of chronic pain, one of the project's aims was to explore whether patients expressed preferences for different VR environments. Four categories of VR experience were devised, and VR applications were purposefully selected and tested in advance for potential efficacy by the researchers for use in each of the 4 weeks of the study. In week 1, the participants undertook passive VR experiences where they simply travelled through a VR environment. These included a virtual Iceland, and a boat ride through an artistic experience (Senza Peso). In the second week, mindfulness and meditative introversion focused VR applications were used, as these have been associated with pain control in other studies [29,30]. These experiences involved flying through 3D mandalas, or experiences that altered the user's environment depending where they looked (Sightline). In the third week, active exploratory VR environments were used, where the participant could explore a new environment at will (an underwater environment, the solar system, and a natural environment). In the final week, active problem-solving experiences were used (eg, game type environments requiring participants to solve 3D puzzles). These different applications allowed for comparison of the VR environments in terms of any reported specific effects on patient's pain experiences and side effects, and also prevented boredom with the VR experiences available at the time. An identical protocol of specific VR experiences to be used for 30 min on every other day of the study was given to each patient (see Figure 1). A simple computer menu system was devised so that patients could easily start each VR application on the appropriate day in sequence (as per protocol). Daily VR activity was logged, with arrest from VR at weekends.

Three sets of equipment were used concurrently with different patients, and then moved on to the next patients who had volunteered on a previously arranged schedule. The equipment was installed in the patients' home with a 90-min training session on how to use the VR and perform data collection procedures. Patients were also assessed as being able to navigate the VR experiences and use the equipment comfortably at the end of this. During the study time patients also had access to a member of the research team by telephone and email for trouble shooting and to discuss any issues associated with the study.

Given an absence of prior work with VR and chronic pain, an initial exposure to 12 therapeutic sessions of 30 min was identified as reasonable to explore the clinical effects of VR initially. VR research in acute pain settings was usually of 15-30 min duration [2] and in associated pain hypnotherapy studies that used 4-6 sessions had less success whereas studies that had offered 8-12 sessions with lengths of treatment exposure between 30 and 40 min established positive results [31,32].

Figure 1. VR (virtual reality) intervention protocol and sequence.



Instruments: Quantitative Tools

To ensure the multidimensional aspects of chronic pain experienced by individuals were adequately measured, various tools were selected that addressed different aspects of chronic pain. These were self-recorded by patients in a supplied binder.

Pre- and Postexposure NRS Scores

Participants were asked to self-rate and record their pain intensity in a diary using the NRS immediately before and after the intervention and at 6 and 24 h postintervention (it was impractical to record pain during the VR experience as this would have proved disruptive to the experience). These measurement points supported the capture of any residual therapeutic effects or trends following the intervention. The NRS meets the IMMPACT group recommendations for using a global impression of change question [33,34].

Weekly Pain Trends: The Brief Pain Inventory (BPI) and Short Leeds Assessment of Neuropathic Symptoms and Signs (S-LANSS)

The BPI and the S-LANSS were used initially (as a baseline) and then at the end of each week of therapy to capture more detailed pain assessment data (giving scores recorded at 5 different time points). The BPI is a validated tool that has been used in numerous studies investigating pain [35-37]. The S-LANSS was specifically designed to measure the pain qualities associated with neuropathic pain and treatment effects. It is a quick to use self-reporting scale consisting of 12 distinct

questions, which ask about the intensity and quality of the patients' pain [38].

Cybersickness Reporting Form

Cybersickness is a well-documented side-effect of VR experiences [28,39-41]. It is the tendency for some users to display symptoms analogous to motion sickness both during and after the VR experience. It is distinct from motion sickness, in that the user is normally stationary but has a compelling sense of self-motion through moving visual imagery. A simple guide to avoid cybersickness and self-reporting form was provided for patients to record any episodes on a weekly basis, indicating the experiences that had led to the episode, onset and duration, any factors that alleviated the sickness, and also any departures from the protocol provided.

Instruments: Qualitative Tools

Initial Individual Interviews

A short semistructured personal interview with each patient was executed at the start of the study by a researcher to capture basic biographic information from participants and a pain history (eg, ethnicity, age, gender, and prior VR and computer gaming experience; pain history: cause, duration, onset, nature, treatment history, and current pharmacological and other interventions used for pain).

Terminal Individual Interviews

An audiotaped 30-min semistructured personal interview was undertaken at the end of the study by a researcher during the final visit to collect the VR equipment. The interview was designed to capture the patient's perceptions of the value of the VR interventions in the management of their pain, their overall impressions of the experience, and any adverse effects.

Data Analysis

Quantitative data were explored for any differences in the average pain scores for the VR intervention at each time point to analyze for any indications of changes in pain immediately following the VR experience, and at 6 and 24 h after exposure, and for cybersickness. Descriptive univariate statistics were analyzed to support a preliminary understanding of the impact of VR on an individual's pain experience and nature of the data obtained. A simple initial pre-post exposure NRS analysis was then undertaken using a Wilcoxon matched pairs test. For the BPI and S-LANSS scores, a Freidman test (the nonparametric alternative to the one-way repeated measure ANOVA) was undertaken to explore for any trends evident over the whole month using SPSS 23 statistical software (IBM).

Qualitative data were transcribed from the original sources into Nvivo 7.0 qualitative analysis software. It was analyzed using an interpretive-description (ID) approach for an open exploration of participant's experiences to further understand the perceptions associated with the use of VR and any impact on their chronic pain. ID assumes preexisting theoretical knowledge, and that clinical patterns exist, and rather than trying to avoid preconceptions in the analysis, coding proceeds on the basis that no matter how participatory and collaborative the analytical method is, it will finally be the researcher who determines what data are significant [42,43]. An ID approach allowed for an inductive descriptive analysis of the phenomena, using iterative readings of the combined qualitative data and coding by 2 independent members of the research team. Analyses were then merged to establish key thematic elements, patterns, and theory associated with the patient's experiences [42].

Results

Sample Characteristics

As is common with chronic pain studies, the study encountered attrition and 8 patients (n=8, 33% attrition) completed the full study protocol, and only 6 of these consented to post experience interviews [44,45]. Reasons for discontinuation were not required, although one indicated it was due to cybersickness. The patients' pain conditions and histories are summarized in Table 1.

Table 1. Participant characteristics. All participants were unemployed, had high-school graduate education levels, and good computer literacy.

ID	Age	Gender	Pain diagnosis	Pain treatment history
01	48	Female	Low back and knee pain following traumatic injury 6 years ago.	Pharmacological: Acetaminophen & Codeine (Tylenol 3), Oxycodone & Acetaminophen (Percocet), Gabapentin, Ibuprofen, Diclofenac, Oxycodone.
			Reported daily NRS score: 4-6	Surgical: Right knee replacement.
				Other: Physiotherapy, Occupational therapy, Intramuscular Stimulation (IMS)
02	63	Female	Arachnoiditis and low back pain following traumatic injury 4 years ago.	Pharmacological: Acetaminophen & Codeine (Tylenol 3), Ibuprophen, Pregabalin, Gabapentin, Cortisone injection, Ketamine, Fentanyl patches, Hydromorphone.
			Reported daily NRS score: 4-7	Surgical: Microdiscectomy, Nerve block.
				Other: Physiotherapy, Chiropractic
04	66	Male	Ilioinguinal neuralgia following hernia repair 14 years ago. Ankylosing Spondylitis over the last 6 years.	Pharmacological: Acetaminophen & Codeine (Tylenol 3), Ibuprophen, Pregabalin, Gabapentin, Ketamine, Methadone, Lignocaine (topical), Oxycodone.
			Reported daily NRS score: 4-8	Surgical: Inguinal surgical mesh removal and inguinal neurectomy.
				Other: Physiotherapy, Yoga
05	50	Male	Cervical spine and shoulder pain following traumatic injury 20 years ago.	Pharmacological: Gabapentin, Clonazepam, Nabilone, Oxycodone, Magnesium Injection, Buprenorphine patches.
			Reported daily NRS score: 5-8	Other: Physiotherapy, Occupational Therapy, Massage, Acupuncture, Water Therapy, Myofascial Release (MFR)
08	71	Female	Chronic hip & lower back pain for 20 years from strain caused through professionally playing classical guitar for 30 years.	Pharmacological: Acetaminophen & Codeine (Tylenol 3), Ibuprophen, Pregabalin.
			Reported daily NRS score: 4-8	Other: Meditation, Naturopathy (prescribed Turmeric)
09	31	Female	Complex regional pain syndrome (Type 2) secondary to thrombosis and multiple embolism 6 years ago	Pharmacological: Acetaminophen & Codeine (Tylenol 3), Naproxen, Cyclobenzaprine, Gabapentin, Hydromorphone, Pregabalin, Fentanyl patches.
			Reported daily NRS score: 6-9	Other: Physiotherapy, Massage Therapy, Intramuscular Stimulation (IMS)
10	43	Female	Migraine headaches and small fiber myopathy following traumatic injury and resulting brain lesion when 7 years old.	Pharmacological: Topiramate, Pregabalin, Cannabis vaporizer and Buprenorphine patch.
			Reported daily NRS score: 5-7	Other: Acupuncture, Chiropractic, Massage therapy
11	36	Female	Low back pain and myofascial pain following traumatic injury 8 years ago	Pharmacological: Acetaminophen & Codeine (Tylenol 3), Ibuprofen, Baclofen, Lidocaine/Ketamine cream, Gabapentin, Pregabalin, Lidocaine injection, Tramadol Reported daily NRS score: 5-8 & Acetaminophen (Tramacet).
				Other: Physiotherapy, Occupational Therapy, Yoga, TENS

Quantitative Analysis

A descriptive statistical exploration of the various pain scores confirmed the data were not normally distributed. Univariate comparison of the mean NRS pre-post test scores demonstrated a slight decrease in pain for most interventions (see Figure 2), although for one intervention (#10) a slight increase in pain was also observed. This likely reflected user frustration with the Subnautica prerelease app, which had not yet implemented game-controller functionality, so participants had to move around using keyboard controls, which was difficult with an HMD on. No reduction in NRS scores 6 and 24 h later was evident. Furthermore, a Wilcoxon Matched Pairs Signed Ranks test for means for each VR intervention demonstrated no significant effect between pre- and postexposure for NRS scores (see Table 2). Given the lack of any significant pre-post

exposure effects, no further exploration of NRS scores was performed. The Friedman test run on the BPI and SLANNS scores as repeated measures for each participant at the end of each week also indicated no statistically significant difference in the pain reported over the 4 weeks. BPI Worst Pain: x^2=1.6, *P*=.82 BPI Average Pain: x^2=5.2, *P*=.27; BPI Least Pain: x^2=4.6, *P*=.20. BPI Pain Now: x^2=2.9, *P*=.57. S-LANSS x^2= 1.0, *P*=.91. As no significant findings were obtained, no further post hoc analysis was performed.

Table 2. NRS scores pre-post VR exposure Wilcoxon signed ranks test.

Value	Post1 - Pre1	Post2 - Pre2	Post3 - Pre3	Post4 - Pre4	Post5 - Pre5	Post6 - Pre6	Post7 - Pre7	Post8 - Pre8	Post9 - Pre9	Post10 - Pre10	Post11 - Pre11	Post12 - Pre12
Z	.000[a]	−1.414[b]	−.322[b]	−1.633b	.000[a]	−1.841[b]	.000[a]	−.680[b]	−1.289[b]	.000[a]	−.816[b]	−.816[b]
Asymptotic Significance (2-tailed)	>.99	.16	.75	.10	>.99	.07	>.99	.50	.20	>.99	.41	.41

[a]The sum of negative ranks equals the sum of positive ranks.

[b]Based on positive ranks.

Qualitative Analysis

Four major thematic areas emerged from the interview analysis: design of the VR experiences, efficacy of VR for chronic pain, limits of the VR technology, and practicality of use as an adjunctive therapy (see Table 3). The subthemes evident in these are described below together with participant quotes.

There was a distinct difference in participants' perceptions of the value of VR environments designed to be interactive versus those designed to promote relaxation. Half of the interviewees believed that the interactive experiences were more beneficial. For example:

> *The space one too, when you're exploring and you're driving, I loved that one too. I thought that was so much fun, I love the ones when you using your brain, and you're actually trying to do stuff, I enjoyed those more, but when I was done, I didn't even notice.* [Participant #10]

> *…because then you are actually immersed in it, whereas some of the ones you felt like even though it was 3D or whatever, you weren't really immersed yeah you were a recipient of the experience, but the ones that made you think and do and that and react were more immersive and then the more immersive it is the more it worked.* [Participant #1]

Two participants reported they actively disliked the relaxation-based VR experiences:

> *I didn't feel that um, I didn't sense that just floating over Iceland or um, some of those other things didn't do it for me* [Participant #4]

Figure 2. Pre-postexposure NRS score differentials by intervention. NRS: Numerical Rating Scale.

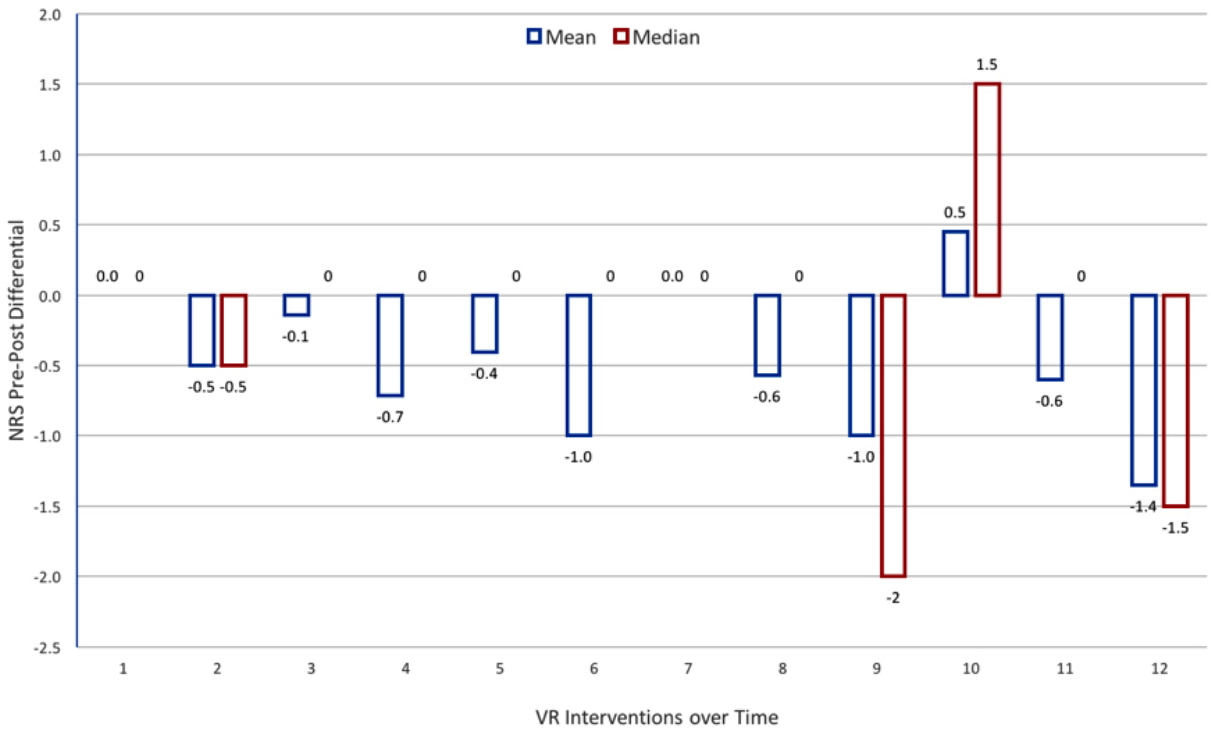


Table 3. Key themes emerging from Interviews (numbers=number of separate responses).

Major theme	Subtheme (positive)	Subtheme (negative)
VR design	Relaxation (7)	Relaxation anxiety (3)
	Interaction (17)	
	Immersion (3)	
	Variety (6)	
	Enjoyment (4)	
Efficacy	Effective (14)	Ineffective (4)
	Distraction (8)	Transient (10)
	Mobility (2)	Stressful (3)
Technological functionality	Comfort (1)	Frustration (13)
		Technical issues (12)
		Graphics quality (4)
		Comfort (2)
Practicality	-	Side effects:
		Cybersickness (6)
		Claustrophobia (1)
		Engagement (5)
		Position (3)

However, 3 participants favored the relaxing experiences as in:

> *…some programs were really good like the guided meditation one I found really good, managed to make me feel really relaxed by the end of it.* [Participant #9]

There appeared no relationship between the type of VR experience people preferred and types of pain they experienced.

The second theme related to the efficacy of the VR experiences in helping reduce pain. Four of those interviewed identified positive benefits during the VR experience:

> *when I was doing them I didn't notice the pain until I was done. So, when I play the games, no pain. But when I filled-in the questionnaire, that's when I noticed it!* [participants #1, 4, 8, and 10]

Participants who experienced positive effects, distraction was frequently highlighted as the rationale:

> *So overall as an experience, if the idea is a distraction to the pain, yeah it worked for a little while then.* [Participant #4]
>
> *…actually take your mind off the world, cause you're in a totally different place, a fantasy.* [Participant #10]

All the participants who found benefits, all reported the experience was transient and did not persist long after the VR experience:

> *The effect on my pain wore off and I was aware of my pain certainly within a few hours of using the VR.* [Participant #1]

Two participants reported the VR experiences were not effective for helping reduce their pain:

> *Overall I would just say it was really interesting, versus like actually helping with my pain.* [Participant #11]

One participant (#10) who had reported positive effects with some of the VR experiences noted one actually caused her stress, increasing her pain:

> *Where I was under water in the ocean, I saw the shark, and I thought it was going to attack me. It was giving me anxiety a bit, and stressing me out a bit, so as soon as I saw that my hand started burning, and my feet…* [Participant 10]

Although conversely, another acknowledged:

> *Some were more stressful…but I find that - when I get stress my pain doesn't increase…* [Participant #11]

The technological functionality (and limits) of the available VR experiences also emerged as noteworthy. Four participants identified that they became quite frustrated with the VR systems in use, because of complex or cumbersome control systems:

> *I had some trouble figuring out which controls to use to move around so um I've never played computer games before and maybe that had something to do with it. I felt like a total idiot totally frustrated and not able to catch onto what to do.* [Participant #8]

The comfort of the HMD drew both positive and negative comments. One user believed that the HMD was "*…nice and comfortable it was relatively small.*" (Participant #1), whereas 2 others complained about issues trying to use the HMD with eyeglasses:

> *It was a bit uncomfortable working with my glasses.* [Participant #4]

Similarly, another theme on the practicality of using VR was evident. The side effect of cybersickness was significant and reported by 5 of the 8 participants (60%) for a least one VR experience in the self-reporting forms, and it also arose as a topic of interest in the interviews:

> *There was a few times where I had to stop because I felt sick because of how fast I felt I was going.* [Participant #11]

Most of the responses noted mild nausea being induced when experiences involved rapid speed or motion, or moving in 3 dimensions (such as in the spacecraft simulator), which was resolved when they slowed the experience down or had a break. However, 1 participant noted the symptoms persisted for some time after the VR experience:

> *I slowed down so I felt a little less bad and I thought I could continue. But afterwards I went and my mom got me a ginger ale, and I laid down and I thought, thank goodness I don't have that in the car or whatever.* [Participant #11]

Two participants also noted minor claustrophobia as a side effect:

> *I felt a bit of claustrophobia because when I was under water and I realized at first I didn't know how to get above the water and was running out of air.* [Participant #4]

The issue of being able to engage with a VR experience when the participant was experiencing severe pain also arose as a practical limitation. In 2 cases, they reported they were in too much pain to use the VR equipment.

Discussion

Principal Findings

Although there were no significant pre-post exposure changes in the reported pain scores, more than half (5 of 8) of the participants did report positive benefits on their perceived pain from the use of VR. However, 3 of the participants interviewed reported none. The effects of VR on chronic pain would appear to be very individualized. No evidence that any benefits of using the VR on the participant's pain persisted postexposure was found. For participants who identified positive results, distraction was described as the mode of action by them. This is consistent with other researchers who suggest the deeper form of distraction produced with VR experiences is the main mechanism by which pain is attenuated [19,20,22,31,46].

Chronic pain patients respond in very individualistic ways to VR as indicated by the varying preferences for interactive versus relaxing forms of VR reported. Some had very negative reactions to the relaxation introversion-focused experiences, whereas others enjoyed them. This finding is consistent with work that reports some individuals are actually relaxation-sensitive and paradoxically find relaxing experiences increase their stress levels [47,48]. Although no significant improvement in BPI and functionality was evident over time, one of the participants did report improved mobility following the VR experiences. Functional improvement has also been reported in other VR studies [49-51].

It was evident that VR technology remains immature in the technology life cycle, the progression from research and development, commercial production, to succession by superior technologies [52,53]. Most participants reported some technical issues with either the hardware or software during the experience. However, the possibility of achieving nonpharmacological pain relief may encourage more chronic pain suffering to experiment with VR, as they are more likely to experiment when pharmacological and other medical therapies fail [54,55].

The most significant adverse reaction to VR was cybersickness, as 60% of the participants experienced this at some time during the study. This side effect of VR is well documented, but should not be underestimated as a factor that may influence uptake [28,39-41,56-59]. Other than cybersickness, no significant adverse effects were noted.

Limitations

As an initial exploratory study, this work has limitations. Case series are vulnerable to selection bias and may not represent the wider population, and with small-scale studies such as this, the effects seen may be due to intervening effects such as the placebo, Hawthorne, or Rosenthal effect. Therefore, internal validity and reliability may be limited. Technological immaturity of the experimental setting is also a limitation. Strengths of the study include the inductive ID approach, which is more responsive to clinical experience-based questions here. Furthermore, an exploratory pilot study provides an appropriate approach at this fundamental stage of clinical research to inform future work.

Work in this area is in its infancy and clinical studies limited. The complexities of chronic pain make finding pain-management solutions challenging, and the results reflect those complexities. The following conclusions were drawn from the results:

Exposure to VR for 30 min a day every other day for chronic pain patients in self-administered therapy sessions resulted in:

- 66% of participants reporting a reduction in pain while using the VR therapy,
- No significant pre- or postexposure differential in pain scores,
- No significant postexposure impact on pain levels,
- No significant postexposure impact on pain interfering with daily function,
- 60% of patients reported episodes of cybersickness when using VR.

The majority of this study participants reported a reduction in pain while they were using the VR, but with highly individualized responses. Findings, as with other recent work suggests that VR maybe a useful short-term adjunct for the management of chronic pain, but individual choice in the form of VR experience may be as significant as the VR medium itself [60]. Although statistically significant reductions in pain scores

postexposure were not demonstrated, in the qualitative analysis, participants reported mostly positive impacts on their pain experience, and one reported that the VR experience did appear to improve their mobility. Whereas longer-term benefits of VR therapy as an adjunctive for chronic pain were not demonstrated, the immediate relief experienced by patients here during VR would indicate the therapy has potential for a means of providing respite from the constant pain they experience.

Conclusions

Attention to practical implementation is important, particularly having good orientation practices and technical support available to patients at home. Robust controlled trials with larger samples, comparing VR with other forms of multimedia and neurological studies are now required to establish efficacy. Practical methods to research VR interventions should include both active and passive interventions, and larger cohort studies including assessment of cybersickness (and factors that ameliorate its effects) as the most significant adverse effect.

In conclusion, home-based VR therapy is a feasible option for chronic pain sufferers. There remains a pressing need for non-opioid alternatives in the treatment of chronic pain, and in light of the patient's experiences documented here, individual tailored VR solutions would appear more likely to be successful compared with a unidimensional off-the-shelf VR experience.

Acknowledgments

The Helen Shore Nursing Endowment Fund is acknowledged for providing the funding to conduct the study. Pain BC and the BC People in Pain Network are acknowledged for assisting in the recruitment of patients for the study. Unknown Worlds is acknowledged for supporting the work by providing their Subnautica VR software at no cost for the research, but was not involved in the data analysis, data interpretation, or in the preparation of this manuscript.

Conflicts of Interest

None declared.

References

1. Rheingold H. Virtual Reality. London: Simon & Schuster; 1991:24-28.
2. Garrett B, Taverner T, Masinde W, Gromala D, Shaw C, Negraeff M. A rapid evidence assessment of immersive virtual reality as an adjunct therapy in acute pain management in clinical practice. Clin J Pain 2014 Dec;30(12):1089-1098. [doi: 10.1097/AJP.0000000000000064] [Medline: 24535053]
3. Guo C, Deng H, Yang J. Effect of virtual reality distraction on pain among patients with hand injury undergoing dressing change. J Clin Nurs 2015 Jan;24(1-2):115-120. [doi: 10.1111/jocn.12626] [Medline: 24899241]
4. Maani CV, Hoffman HG, Morrow M, Maiers A, Gaylord K, McGhee LL, et al. Virtual reality pain control during burn wound debridement of combat-related burn injuries using robot-like arm mounted VR goggles. J Trauma 2011 Jul;71(1 Suppl):S125-S130 [FREE Full text] [doi: 10.1097/TA.0b013e31822192e2] [Medline: 21795888]
5. JahaniShoorab N, Ebrahimzadeh Z, Nahvi A, Mazluom S, Golmakani N, Talebi M, et al. The effect of virtual reality on pain in primiparity women during episiotomy repair: a randomize clinical trial. Iran J Med Sci 2015 May;40(3):219-224 [FREE Full text] [Medline: 25999621]
6. Villiger M, Bohli D, Kiper D, Pyk P, Spillmann J, Meilick B, et al. Virtual reality-augmented neurorehabilitation improves motor function and reduces neuropathic pain in patients with incomplete spinal cord injury. Neurorehabil Neural Repair 2013 Oct;27(8):675-683. [doi: 10.1177/1545968313490999] [Medline: 23757298]
7. Walker MR, Kallingal GJ, Musser JE, Folen R, Stetz MC, Clark JY. Treatment efficacy of virtual reality distraction in the reduction of pain and anxiety during cystoscopy. Mil Med 2014 Aug;179(8):891-896. [doi: 10.7205/MILMED-D-13-00343] [Medline: 25102532]
8. Jones T, Moore T, Rose H, Choo J. The impact of virtual reality on chronic pain. J Pain 2016 Apr;17(4S):S102-S103. [doi: 10.1016/j.jpain.2016.01.319] [Medline: 28162315]
9. Botella C, Garcia-Palacios A, VizYcaíno Y, Herrero R, Baños RM, Belmonte M. Virtual reality in the treatment of fibromyalgia: a pilot study. Cyberpsychol Behav Soc Netw 2013 Mar;16(3):215-223. [doi: 10.1089/cyber.2012.1572] [Medline: 23496678]
10. Ramachandran V, Seckel E. Using mirror visual feedback and virtual reality to treat fibromyalgia. Med Hypotheses 2010 Dec;75(6):495-496. [doi: 10.1016/j.mehy.2010.07.003] [Medline: 20692106]
11. Schopflocher D, Taenzer P, Jovey R. The prevalence of chronic pain in Canada. Pain Res Manag 2011;16(6):445-450 [FREE Full text] [Medline: 22184555]
12. Smith BH, Elliott AM, Hannaford PC, Royal College of General Practitioners' Oral Contraception Study. Is chronic pain a distinct diagnosis in primary care? Evidence arising from the Royal College of General Practitioners' Oral Contraception study. Fam Pract 2004 Feb;21(1):66-74. [Medline: 14760048]
13. IASP. IASP. 2012. International Association for the Study of Pain (IASP) Taxonomy URL: http://www.iasp-pain.org/Taxonomy?navItemNumber=576 [accessed 2017-04-14] [WebCite Cache ID 6pivXrIGO]

14. Canadian Pain Coalition (CPC). 2014. The Painful Truth: The state of pain management in Canada URL: http://www.canadianpaincoalition.ca/media/painful_truth_report.pdf [accessed 2017-04-14] [WebCite Cache ID 6pivbgLHg]
15. Rashiq S, Schopflocher D, Taenzer P, Jonsson E. The Economics of Chronic Pain. In: Chronic Pain. Weinheim, Germany: Wiley-Blackwell; Sep 2008:41-56.
16. Smith BH, Elliott AM, Hannaford PC, Royal College of General Practitioners' Oral Contraception Study. Is chronic pain a distinct diagnosis in primary care? Evidence arising from the Royal College of General Practitioners' Oral Contraception study. Fam Pract 2004 Feb;21(1):66-74. [Medline: 14760048]
17. Gilmour H. Chronic pain, activity restriction and flourishing mental health. Health Rep 2015 Jan;26(1):15-22 [FREE Full text] [Medline: 25606984]
18. Gatchel R, Peng Y, Peters M, Fuchs P, Turk D. The biopsychosocial approach to chronic pain: scientific advances and future directions. Psychol Bull 2007 Jul;133(4):581-624. [doi: 10.1037/0033-2909.133.4.581] [Medline: 17592957]
19. Sander WS, Eshelman D, Steele J, Guzzetta CE. Effects of distraction using virtual reality glasses during lumbar punctures in adolescents with cancer. Oncol Nurs Forum 2002;29(1):E8-E15. [doi: 10.1188/02.ONF.E8-E15] [Medline: 11845217]
20. Schneider S, Prince-Paul M, Allen M, Silverman P, Talaba D. Virtual reality as a distraction intervention for women receiving chemotherapy. Oncol Nurs Forum 2004;31(1):81-88. [doi: 10.1188/04.ONF.81-88] [Medline: 14722591]
21. Melzack R, Wall PD. Pain mechanisms: a new theory. Science 1965 Nov 19;150(3699):971-979. [Medline: 5320816]
22. McCaul KD, Malott JM. Distraction and coping with pain. Psychol Bull 1984 May;95(3):516-533. [Medline: 6399756]
23. Melzack R. Evolution of the neuromatrix theory of pain. 2004 May Presented at: The Prithvi Raj Lecture: The third World Congress of the World Institute of Pain; 2004; Barcelona, Spain p. 85-94.
24. Bantick SJ, Wise RG, Ploghaus A, Clare S, Smith SM, Tracey I. Imaging how attention modulates pain in humans using functional MRI. Brain 2002 Feb;125(Pt 2):310-319. [Medline: 11844731]
25. Gold J, Belmont K, Thomas D. The neurobiology of virtual reality pain attenuation. Cyberpsychol Behav 2007 Aug;10(4):536-544. [doi: 10.1089/cpb.2007.9993] [Medline: 17711362]
26. Mura G. The Meditation Chamber. In: Metaplasticity in Virtual Worlds: Aesthetics and Semantic Concepts. Hershey, PA: IGI Global; 2010:133-156.
27. Shaw C, Gromala D, Seay AF. The meditation chamber: enacting autonomic senses. 2007 Nov Presented at: Enactive 07; November 19-22, 2007; Grenoble, France URL: http://www.sfu.ca/~shaw/papers/Enactive07MedChamber.pdf
28. LaViola J. A discussion of cyber-sickness in virtual environments. SIGCHI Bulletin 2000;32(1):47-55 [FREE Full text]
29. Zeidan F, Vago DR. Mindfulness meditation-based pain relief: a mechanistic account. Ann N Y Acad Sci 2016 Jun;1373(1):114-127. [doi: 10.1111/nyas.13153] [Medline: 27398643]
30. de Boer MJ, Steinhagen HE, Versteegen GJ, Struys MM, Sanderman R. Mindfulness, acceptance and catastrophizing in chronic pain. PLoS One 2014;9(1):e87445 [FREE Full text] [doi: 10.1371/journal.pone.0087445] [Medline: 24489915]
31. Soltani M, Teeley AM, Wiechman SA, Jensen MP, Sharar SR, Patterson DR. Virtual reality hypnosis for pain control in a patient with gluteal hidradenitis: a case report(). Contemp Hypn Integr Ther 2011;28(2):142-147 [FREE Full text] [Medline: 23205274]
32. Elkins G, Johnson A, Fisher W. Cognitive hypnotherapy for pain management. Am J Clin Hypn 2012 Apr;54(4):294-310. [doi: 10.1080/00029157.2011.654284] [Medline: 22655332]
33. Breivik EK, Björnsson GA, Skovlund E. A comparison of pain rating scales by sampling from clinical trial data. Clin J Pain 2000 Mar;16(1):22-28. [Medline: 10741815]
34. Herr KA, Spratt K, Mobily PR, Richardson G. Pain intensity assessment in older adults: use of experimental pain to compare psychometric properties and usability of selected pain scales with younger adults. Clin J Pain 2004;20(4):207-219. [Medline: 15218405]
35. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med Singapore 1994 Mar;23(2):129-138. [Medline: 8080219]
36. Sawatzky BJ, Slobogean GP, Reilly CW, Chambers CT, Hol AT. Prevalence of shoulder pain in adult- versus childhood-onset wheelchair users: a pilot study. J Rehabil Res Dev 2005;42(3 Suppl 1):1-8 [FREE Full text] [Medline: 16195958]
37. Stroud M, Turner J, Jensen M, Cardenas D. Partner responses to pain behaviors are associated with depression and activity interference among persons with chronic pain and spinal cord injury. J Pain 2006 Feb;7(2):91-99. [doi: 10.1016/j.jpain.2005.08.006] [Medline: 16459274]
38. Bennett MI, Smith BH, Torrance N, Potter J. The S-LANSS score for identifying pain of predominantly neuropathic origin: validation for use in clinical and postal research. J Pain 2005 Mar;6(3):149-158. [doi: 10.1016/j.jpain.2004.11.007] [Medline: 15772908]
39. Bonato F, Bubka A, Palmisano S. Combined pitch and roll and cybersickness in a virtual environment. Aviat Space Environ Med 2009 Nov;80(11):941-945. [Medline: 19911517]
40. Ling Y, Brinkman WP, Nefs HT, Quo C, Heynderickx I. Cyber-sickness anxiety in virtual environments. 2011 Presented at: VTT Symposium (Valtion Teknillinen Tutkimuskeskus); 2011; Delft, Netherlands URL: http://repository.tudelft.nl/islandora/object/uuid:1f5fff10-ce0d-4fef-9c74-896f169535e9/datastream/OBJ/view
41. McCauley M, Sharkey T. Cybersickness: perception of self-motion in virtual environments. Presence-Teleop Virt 1992 Jan;1(3):311-318 [FREE Full text] [doi: 10.1162/pres.1992.1.3.311]

42. Thorne S, Kirkham SR, MacDonald-Emes J. Interpretive description: a noncategorical qualitative alternative for developing nursing knowledge. Res Nurs Health 1997 Apr;20(2):169-177. [Medline: 9100747]
43. Thorne S, Sandelowski M. Interpretive Description (Developing Qualitative Inquiry). Walnut Creek, CA: Left Coast Press; 2008.
44. Davis M, Addis M. Predictors of attrition from behavioral medicine treatments. Ann Behav Med 1999;21(4):339-349. [Medline: 10721442]
45. Turk DC, Rudy TE. Neglected factors in chronic pain treatment outcome studies--referral patterns, failure to enter treatment, and attrition. Pain 1990 Oct;43(1):7-25. [Medline: 2277718]
46. Wolitzky K, Fivush R, Zimand E, Hodges L, Rothbaum B. Effectiveness of virtual reality distraction during a painful medical procedure in pediatric oncology patients. Psychol Health 2005 Dec;20(6):817-824 [FREE Full text] [doi: 10.1080/14768320500143339]
47. Heide F, Borkovec T. Relaxation-induced anxiety: mechanisms and theoretical implications. Behav Res Ther 1984;22(1):1-12. [Medline: 6365071]
48. Luberto C. Thesis. Cincinnati, OH: University of Cincinnati; 2012. Development and Validation of the Relaxation Sensitivity Index URL: https://etd.ohiolink.edu/pg_10?215996591039095::NO:10:P10_ETD_SUBID:84440 [accessed 2017-04-14] [WebCite Cache ID 6pj07Pakr]
49. Cole J, Crowle S, Austwick G, Slater D. Exploratory findings with virtual reality for phantom limb pain; from stump motion to agency and analgesia. Disabil Rehabil 2009;31(10):846-854. [doi: 10.1080/09638280802355197] [Medline: 19191061]
50. van Twillert B, Bremer M, Faber A. Computer-generated virtual reality to control pain and anxiety in pediatric and adult burn patients during wound dressing changes. J Burn Care Res 2007;28(5):694-702. [doi: 10.1097/BCR.0B013E318148C96F] [Medline: 17667488]
51. Thomas J, France C, Applegate M, Leitkam S, Pidcoe P, Walkowski S. Effects of visual display on joint excursions used to play virtual dodgeball. JMIR Serious Games 2016 Sep 15;4(2):e16 [FREE Full text] [doi: 10.2196/games.6476] [Medline: 27634561]
52. Bayus B. An analysis of product lifetimes in a technologically dynamic industry. Management Sci 1998;44(6):775 [FREE Full text]
53. Rogers E. Diffusion of Innovations, 5th edition. New York, NY: Free Press; 2003.
54. Astin J. Why patients use alternative medicine: results of a national study. JAMA 1998 May 20;279(19):1548-1553. [Medline: 9605899]
55. Thorne S, Paterson B, Russell C, Schultz A. Complementary/alternative medicine in chronic illness as informed self-care decision making. Int J Nurs Stud 2002 Sep;39(7):671-683. [Medline: 12231024]
56. Fernandes A, Feiner S. Combating VR sickness through subtle dynamic field-of-view modification. 2016 Presented at: IEEE Symposium on 3D User Interfaces (3DUI); 2016; Greenville, SC p. 201-210 URL: http://www.cs.columbia.edu/2016/combating-vr-sickness/images/combating-vr-sickness.pdf
57. Kolasinski E. DTIC. 1995. Simulator sickness in virtual environments URL: http://www.dtic.mil/get-tr-doc/pdf?AD=ADA295861 [accessed 2017-04-14] [WebCite Cache ID 6pj1pNE08]
58. Barrett J. Semanticscholar. Edinburgh, Australia: Australian Department of Defense; 2004. Side effects of virtual environments: a review of the literature URL: https://pdfs.semanticscholar.org/ab1b/4153e44abb4c1a1fcac5f2aaee847d30ecf3.pdf [accessed 2017-04-14] [WebCite Cache ID 6pj21DVzj]
59. Merhi O, Faugloire E, Flanagan M, Stoffregen T. Motion sickness, console video games, and head-mounted displays. Hum Factors 2007 Oct;49(5):920-934. [doi: 10.1518/001872007X230262] [Medline: 17915607]
60. Mosadeghi S, Reid M, Martinez B, Rosen B, Spiegel Brennan Mason Ross. Feasibility of an immersive virtual reality intervention for hospitalized patients: an observational cohort study. JMIR Ment Health 2016 Jun 27;3(2):e28 [FREE Full text] [doi: 10.2196/mental.5801] [Medline: 27349654]

Abbreviations

FOV: field of view
HMD: head-mounted display
VR: virtual reality
NRS: Numerical Rating Scale
S-LANSS: Self-completed Leeds Assessment of Neuropathic Symptoms and Signs pain scale

Edited by G Eysenbach; submitted 05.01.17; peer-reviewed by C Lalloo, T Jones; comments to author 23.02.17; revised version received 01.03.17; accepted 18.03.17; published 11.05.17.

Please cite as:
Garrett B, Taverner T, McDade P
Virtual Reality as an Adjunct Home Therapy in Chronic Pain Management: An Exploratory Study
JMIR Med Inform 2017;5(2):e11
URL: http://medinform.jmir.org/2017/2/e11/
doi:10.2196/medinform.7271
PMID:28495661

Original Paper

Validation of an Improved Computer-Assisted Technique for Mining Free-Text Electronic Medical Records

Marco Duz[1], MedVet, MVM, DipECEIM, PhD, MRCVS; John F Marshall[2], BVMS, PhD, DipACVS, DipECVS, MRCVS; Tim Parkin[2], BSc, BVSc, PhD, dipECVPH, FHEA, MRCVS

[1]School of Veterinary Medicine and Science, University of Nottingham, Loughborough, United Kingdom

[2]School of Veterinary Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom

Corresponding Author:
Marco Duz, MedVet, MVM, DipECEIM, PhD, MRCVS
School of Veterinary Medicine and Science
University of Nottingham
Sutton Bonington Campus
College Road
Loughborough, LE12 5RD
United Kingdom
Phone: 44 115 951 6796
Fax: 44 1159516415
Email: marco.duz@nottingham.ac.uk

Abstract

Background: The use of electronic medical records (EMRs) offers opportunity for clinical epidemiological research. With large EMR databases, automated analysis processes are necessary but require thorough validation before they can be routinely used.

Objective: The aim of this study was to validate a computer-assisted technique using commercially available content analysis software (SimStat-WordStat v.6 (SS/WS), Provalis Research) for mining free-text EMRs.

Methods: The dataset used for the validation process included life-long EMRs from 335 patients (17,563 rows of data), selected at random from a larger dataset (141,543 patients, ~2.6 million rows of data) and obtained from 10 equine veterinary practices in the United Kingdom. The ability of the computer-assisted technique to detect rows of data (cases) of colic, renal failure, right dorsal colitis, and non-steroidal anti-inflammatory drug (NSAID) use in the population was compared with manual classification. The first step of the computer-assisted analysis process was the definition of inclusion dictionaries to identify cases, including terms identifying a condition of interest. Words in inclusion dictionaries were selected from the list of all words in the dataset obtained in SS/WS. The second step consisted of defining an exclusion dictionary, including combinations of words to remove cases erroneously classified by the inclusion dictionary alone. The third step was the definition of a reinclusion dictionary to reinclude cases that had been erroneously classified by the exclusion dictionary. Finally, cases obtained by the exclusion dictionary were removed from cases obtained by the inclusion dictionary, and cases from the reinclusion dictionary were subsequently reincluded using Rv3.0.2 (R Foundation for Statistical Computing, Vienna, Austria). Manual analysis was performed as a separate process by a single experienced clinician reading through the dataset once and classifying each row of data based on the interpretation of the free-text notes. Validation was performed by comparison of the computer-assisted method with manual analysis, which was used as the gold standard. Sensitivity, specificity, negative predictive values (NPVs), positive predictive values (PPVs), and F values of the computer-assisted process were calculated by comparing them with the manual classification.

Results: Lowest sensitivity, specificity, PPVs, NPVs, and F values were 99.82% (1128/1130), 99.88% (16410/16429), 94.6% (223/239), 100.00% (16410/16412), and 99.0% (100×2×0.983×0.998/[0.983+0.998]), respectively. The computer-assisted process required few seconds to run, although an estimated 30 h were required for dictionary creation. Manual classification required approximately 80 man-hours.

Conclusions: The critical step in this work is the creation of accurate and inclusive dictionaries to ensure that no potential cases are missed. It is significantly easier to remove false positive terms from a SS/WS selected subset of a large database than search that original database for potential false negatives. The benefits of using this method are proportional to the size of the dataset to be analyzed.

(JMIR Med Inform 2017;5(2):e17) doi:10.2196/medinform.7123

KEYWORDS

text mining; data mining; electronic medical record; validation studies

Introduction

Exploitation of clinical information in electronic medical records (EMRs) has the potential to revolutionize medical research. Even though time consuming, the assumed gold standard for analysis of medical data consists of manual evaluation, and this is what automated analysis tools should be validated against [1]. Data in an EMR could be used to perform epidemiological studies to support updated disease registries, drug safety surveillance, clinical trials, and health audits [2]. One of the aims of the Department of Health in the United Kingdom is to achieve a paperless National Health Service (NHS) by 2018 [3]. This plan would result in conversion of the medical records of the whole British population to a digital format, and therefore, potentially make it available for epidemiological research. Efficiency of algorithms for anonymization of an EMR have also been thoroughly evaluated [4,5], so considerations related to protection of patient's confidentiality are unlikely to pose a limitation to these studies. Many EMR management systems are currently in use in medical practice, and this poses a challenge to research as these systems store data using different formats. However, while these systems present substantial technical differences, at a minimum, data is generally stored with a combination of structured data (patient ID, location, and date) and unstructured free-text clinical notes, often including further information such as diagnostic imaging, laboratory reports, and billing information where applicable. Data stored in these systems could be used for epidemiologic research using methodologies that are independent from the system used [2]. Coding of medical records is often implemented to support a clear classification of clinical cases but limits a clinician's freedom of expression and often does not entirely suit all details of the clinical case and relies on clinicians to use the coding system correctly [6].

Text mining techniques have been developed over the past 40 years [7-10]; they use tools compatible with both structured and unstructured data. Text mining aims to extract information of interest from a dataset and transform this information into an understandable structure for future use [11]. A recent study compared the accuracy of information extraction between the main text mining tools currently available for the purpose of case-detection for named clinical conditions [2]. Commonly used methods include rule-based neuro-linguistic programing (NLP) algorithms that combine basic keyword searching with rules to identify negations or context modifying instances and carry variable sensitivity, specificity, and lower negative predictive values (NPVs) and positive predictive values (PPVs) [2]. A low PPV suggests a poor performance by the algorithm to detect negations and context modifying instances so that sentences that should be excluded from the output search are ultimately included. However, it should be pointed out that PPV is affected by the overall prevalence of the condition of interest. With conditions of low prevalence, a significant proportion of false positives can be identified with anything but close to 100% specificity [2]. Ultimately, any of these algorithms should be tested using conditions of variable prevalence, to describe performance in light of disease characteristics. Recently, lack of standardization in reporting text mining algorithm performance has been described with a suggestion to particularly include data on sensitivity and PPV particularly in these studies [2].

Computer-assisted methodologies to extract information of interest from free-text EMRs have been used to report disease prevalence and for syndromic surveillance in veterinary medicine [12,13] with similar performance to algorithms utilized for processing human EMRs [13]. Although EMR-based research is more limited in veterinary medicine, a method using commercially available software (WordStat, Provalis Research) has been validated showing great potential in EMR-based research, whether veterinary or medical. The advantage of this software is a user-friendly interface that requires minimal training for the operator, and therefore, would be suitable for use by operators without a background in bioinformatics. The software provides the opportunity to adopt user-defined rules to identify negation terms and improve the specificity and PPV [13]. Despite obvious anatomical and pathophysiological differences between human and veterinary patients, EMR management systems share similar structure, goals, and modalities, and the use of text-mining procedures to search data of interest stored as free-text in veterinary EMR databases would also be applicable to human medical EMR databases.

The aim of this study was to describe in detail the text mining process using WordStat and validate its use against manual analysis performed separately by an experienced clinician by reading and interpreting the same data.

Methods

Data Used for Validation

The same dataset (validation dataset) was used for computer-assisted and manual analyses. This was created from a random selection of lifelong clinical records that were extracted using statistical software (R v3.0.2) from a random sample of equine patients from a greater dataset of 2,653,698 rows of data (cases), including 538,193 unique words used a total of 52,039,966 times, from 141,543 patients, obtained from 10 first opinion equine veterinary practices in the United Kingdom, and stored in the .csv format. Each case identifies the content pertinent to that patient on a single row of the .csv file. Each row of data had been generated at each visit but multiple rows could have been generated on that same visit, for example, one row could have been reporting clinical findings, another drug dispensed, and another some management notes. One patient would have contained from a single to several rows of data.

Validation was performed on 4 categories, including three conditions and one for drug use. Colic is a condition of middle-high prevalence in the horse population. Right dorsal colitis and renal failure were included as conditions with a low prevalence in first opinion settings. Finally, a fourth category of non-steroidal anti-inflammatory drug (NSAID) was included to validate the mining process to identify medication prescribing.

Validation Process Design

This study compares the described computer-assisted classification process with that of manual analysis, which is included as the gold standard method of interpretation and classification of free-text clinical notes. The study was completed sequentially in 3 main steps. The first step consisted in the computer-assisted classification. The second step consisted in manual classification. The third step was the comparison of the results from each classification technique. Sensitivity, specificity, PPVs, NPVs, and *F* values of the computer-assisted process compared with the manual process were subsequently calculated by looking at where discrepancies were present between the two classification processes.

Computer-Assisted Classification

Inclusion Dictionary

The first step of the computer-assisted classification consisted in the manual evaluation of the list of all words included in the dataset that had been created by the function "Frequencies" in WordStat and then exported into a spreadsheet. Any word that might have identified any of the above categories was included in the inclusion dictionary relevant for that category. This included terms spelled correctly, spelled incorrectly but judged to likely refer to one of these categories, or abbreviated. Following the creation of the inclusion dictionaries, cases containing words contained in the categorization dictionary were extracted via the "keyword-in-context" function.

The result of the search for each of the categorization dictionaries consisted in a spreadsheet with 5 columns: one for data row number, one for the text preceding the word or combination of words identified by the search, one for the word or combination of words itself, one for the text after the word or combination of words, and one for the patient's anonymous identification. The spreadsheet was saved as a .csv for subsequent evaluation.

Exclusion Dictionary

The output of the search obtained from the inclusion dictionary was subsequently evaluated manually to identify word combinations identifying false positive cases. Each exclusion dictionary was created with combinations of words identifying false positive cases. Once a comprehensive exclusion dictionary had been created, cases containing combinations of words in the exclusion dictionary were identified and were exported as a .csv file.

Both search results from inclusion and exclusion characterization dictionaries were imported in R v3.0.2 so that the rows containing false positive data identified by the exclusion dictionary could be removed from the search results of the inclusion dictionary.

The dataset "Results" included all the cases from the original dataset that included words in the inclusion dictionary but that also excluded the cases containing combinations of words specified in the exclusion dictionary.

Reinclusion Dictionary

The output search of the exclusion dictionary was also evaluated manually to identify whether it included false negative cases. Combinations of words uniquely identifying these false negatives were included in the reinclusion dictionary and reincluded in the Results dataset using R v3.0.2.

The Results datasets obtained for each of the 4 categories investigated by computer-assisted classification consisted in a subset of the validation dataset including the records (with the original row-number) identifying one of the four categories sought. The text-mining process is summarized in Figure 1.

Manual Classification

The validation dataset was classified manually independently from computer-assisted classification. Manual classification was performed entirely in MS Excel where a column for each of the 4 categories investigated was added to the spreadsheet containing the original data. Each row of data was manually tagged according to one of the categories: "NSAIDs," "colic," "renal failure," and "right dorsal colitis." Where a row of data identified more than one category, multiple tags were applied accordingly. Manual classification was performed by a single, experienced equine clinician (holding a degree in veterinary medicine as well as specialist qualification in equine internal medicine), and tag allocation was based on the interpretation of each row of data in the dataset. The dataset was manually evaluated once.

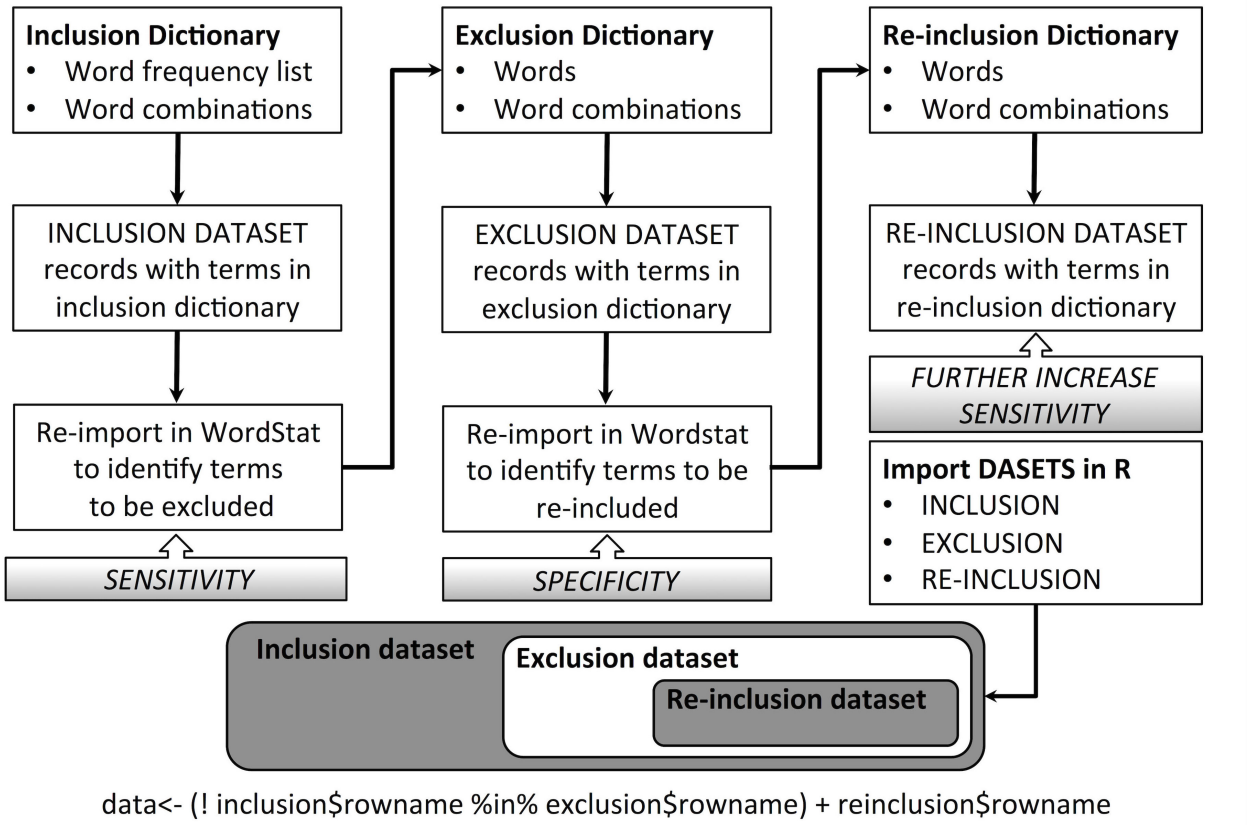
Manual classification results consisted in the .csv validation dataset with an added column for each of the categories sought and a tag in the rows identifying the relevant category.

Comparison between Computer-Assisted and Manual Classifications

The search output for each of the 4 characterization dictionaries and the validation dataset, including the column with the manual classification, were imported in R v3.0.2. Computer-assisted and manual classifications were compared and any discrepancy recorded and subsequently reevaluated manually to investigate the source of the disagreement. A two-by-two contingency table was produced for each dictionary and sensitivity, specificity, PPVs, NPVs, and *F*-measure were calculated.

The study was performed with the approval of the Research Ethics Committee of the School of Veterinary Medicine of the College of Medical, Veterinary and Life Sciences at the University of Glasgow.

Figure 1. Flowchart summary of the text mining process adopted in the study. The lower portion of the picture summarizes how the final dataset (dark gray) resulted from the subtraction of the exclusion dataset from the inclusion dataset and the final addition of reinclusion dataset obtained from the exclusion dataset.



Results

Data Used for Validation

The clinical records of 335 patients, including 17,561 cases from the main dataset, were obtained representing 0.2% of all animals and 0.7% of all data. The average number of cases per animal was 52.3 (median 14; range 1-1031; 1st quartile: 5; 3rd quartile: 44.2). Free-text data included 16,882 unique words used a total of 538,193 times. The data included columns for anonymous patient ID, date of data entry, and a column for free-text clinical notes, which included a mixture of notes entered by the clinician as well as text, including information of drug prescription and sales. There was no standardized diagnostic coding or fixed vocabulary in the dataset.

Computer-Assisted Classification

Inclusion Dictionary

The inclusion dictionary for "NSAIDs" included 53 words, 57 words for "colic," 13 words for "renal failure," and six words for "right dorsal colitis." Words in the NSAID inclusion dictionary were present 1562 times in 1181 cases for NSAIDs, 356 times in 291 cases for colic, 23 times in 23 cases for renal failure, and 7 times in 7 cases for right dorsal colitis.

Exclusion Dictionary

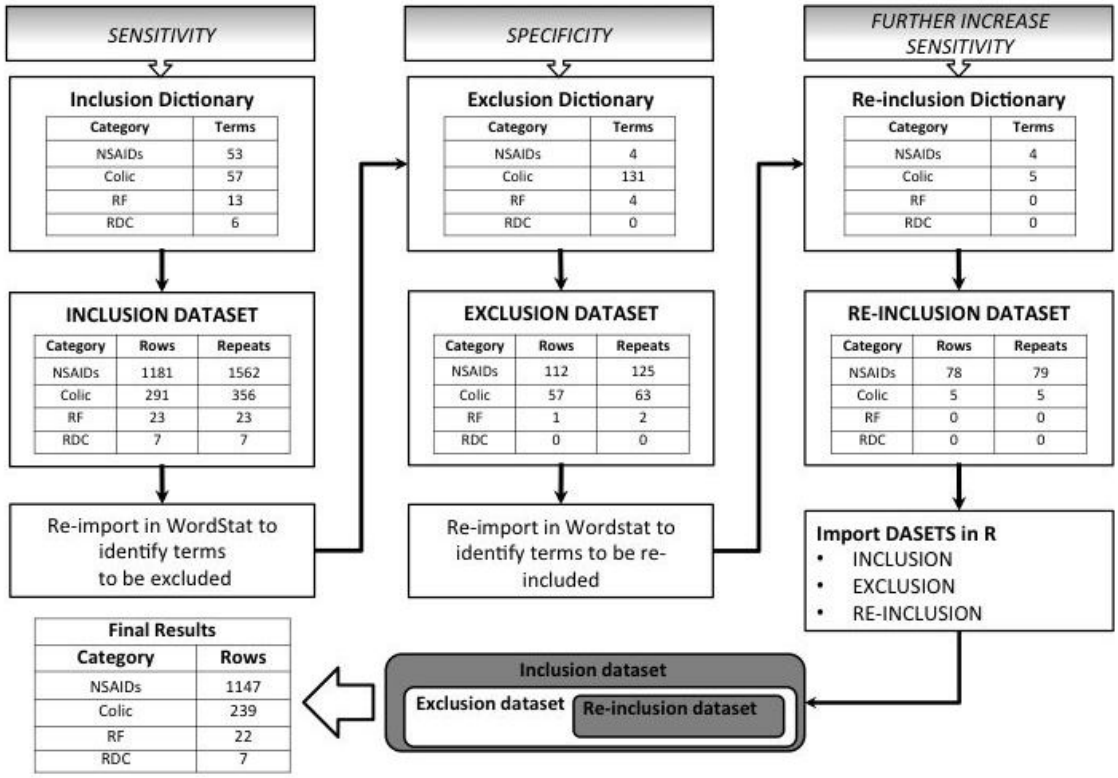
The exclusion dictionary for "NSAIDs" included 4 combinations of words, 131 combinations of words for "colic," 4 combinations of words for "renal failure," and none for "right dorsal colitis." The combinations of words in the NSAIDs exclusion dictionary were present 125 times in 112 cases, 63 times in 57 cases for colic, twice in one case for renal failure, and none for right dorsal colitis.

Reinclusion Dictionary

The reinclusion dictionary for "NSAIDs" included 4 combinations of words, 5 combinations of words for "colic," and none for "renal failure" and "right dorsal colitis." Following data extraction of the total of 17,561 cases in the validation dataset, combinations of words in the NSAIDs reinclusion dictionary were present 79 times in 78 cases and 5 times in 5 cases for colic. No term was present for both renal failure and right dorsal colitis.

Computer-assisted classification was performed in seconds, though the process of dictionary creation was lengthy and required approximately 30 h in total for a dataset of this size. Computer-assisted classification resulted in the identification of 1147 cases for NSAIDs, 239 cases for colic, 22 cases for renal failure, and 7 cases for right dorsal colitis. Data flow for the computer-assisted classification process is summarized in Figure 2.

Figure 2. Flowchart summary of data flow of the computer-assisted text mining process for the validation dataset. The columns include either the number of terms (each term is either a word or a combination of words) in each dictionary, cases, or number of times ("repeats") terms in the relevant dictionary are identified in the dataset.



Manual Classification

Manual classification of the validation dataset of 17,561 cases identified 1130 cases where NSAID prescribing was identified, 226 cases for colic, 22 cases for renal failure, and 7 cases for right dorsal colitis. Manual classification was completed in 80 h performed over a period of 10 days.

Comparison Between Computer-Assisted and Manual Classifications

The results of the comparison of computer-assisted and manual analysis including determination of sensitivity, specificity, PPV, NPV, and *F* value are summarized in Table 1. Overall, there was excellent agreement between computer-assisted and manual analysis.

Computer-assisted classification correctly identified 19 cases that were erroneously classified as negative by manual classification. These were cases where NSAIDs had in fact been prescribed but were missed while reading through the dataset once. Two further cases were false negative cases incorrectly classified by the computer-assisted process but correctly identified by manual processing. These referred to a hypothetical or future use of the drug. For example, one sentence commented that "c/s + NSAID not option (as low TP and previous laminitis)," and the other wrote "NSAID in future." Similarly, there were 13 cases classified as "colic" by the computer-assisted classification but classified not as colic by the manual process. These cases referred to instances where it was not clear whether the case was indeed a colic, where the investigation for colic have been performed but the results are not consistent with colic. For example, "neighbours reported horses colicking now seems fine HR etc normal," the colic was unconfirmed or very short lived so it is debatable whether this really had consisted in a colic case. Another example such as "soft F+ present in rectum.no impaction palpable" suggests that there findings are unremarkable, yet if trans-rectal palpation was performed and noting the lack of an impaction was required, then the horse might indeed have exhibited signs of colic, so again, whether this case should be classified as "colic" is open to debate. The dataset included 22 cases referring to renal failure and 7 cases referring to right dorsal colitis and were all correctly identified by both methodologies.

Table 1. Sensitivity, specificity, and positive and negative predictive values (PPV and NPV, respectively) of computer-assisted analysis compared with manual analysis reported (values reported as per cent values). Rows are conditions identified by the software and columns correspond to manual classification.

Category		Manual[a]		Sensitivity	Specificity	PPV[b]	NPV[c]	*F*
	C-A[d]	+	−					
NSAIDs[e]								
	+	1128	19	99.8	99.9	98.3	100	99.0
	−	2	16410					
Colic								
	+	226	13	100	99.9	94.6	100	100
	−	0	17322					
RF[f]								
	+	22	0	100	100	100	100	100
	−	0	17539					
RDC[g]								
	+	7	0	100	100	100	100	100
	−	0	17554					

[a]+/− in the "manual" column identifies the number of positive and negative terms classified manually in each category (Colic, nonsteroidal antiinflammatory drugs [NSAIDs], renal failure [RF], and right dorsal colitis [RDC]).

[b]PPV: positive predictive values.

[c]NVP: negative predictive values.

[d]The +/− in the "C-A" column identifies the number of positive and negative terms classified with the computer-assisted method for each category.

[e]NSAIDs: nonsteroidal antiinflammatory drugs.

[f]RF: renal failure.

[g]RDC: right dorsal colitis.

Discussion

Principal Findings

The findings of this study show that the methodology described yields results very similar to manual analysis. This methodology is suitable for studies using large free-text EMRs that require the highest possible sensitivity, specificity, PPVs, and NVPs. The technical time required to automatically mine the information of interest from the dataset is negligible (after creation of the relevant dictionaries) in comparison with that of manual analysis. A few seconds are required with the computer-assisted process, depending on machine power, compared with approximately 80 h for the manual analysis for a dataset of 17,561 cases. The study reported here used less than 1% of all available records, and this differential in time required will obviously only get bigger as the dataset increases in size. It is important to point out that a variable amount of time is necessary initially to create adequately comprehensive dictionaries, and this is dependent on the type of dictionary that is being created. Creation of dictionaries identifying a specific diagnosis or to identify drug prescription is faster as the terminology used by clinicians is generally limited and specific. On the other hand, creation of dictionaries that identify a syndrome or a list of generic clinical signs or presenting complaints is highly dependent on the multitude of possible colloquial descriptions that might identify that condition. Creation of exclusion and inclusion dictionaries also requires a degree of manual evaluation of the search output, which would be more time consuming for larger datasets, but not in a linear manner. For example, the number of unique words in the validation dataset was proportionally 10 times greater than in the original dataset. In this study it was noticed that in most cases a relatively small number of word combinations identified the vast majority of false positive and false negative cases, which made exclusion and reinclusion dictionary definition much faster. Furthermore, alphabetic ordering of records by keyword and in context evaluation were performed rapidly as clinicians have the tendency to adopt the same combination of words to describe similar clinical scenarios, hence, reducing grammatical variability and speeding up the analysis process. A novelty of this methodology is in the use of a reinclusion dictionary, which promotes a further increase in the overall method specificity without compromising in sensitivity. This method is therefore suitable for studies where optimal identification of cases is required. A further advantage is the ease-of-use of the software that makes this method suitable to operators without any prior background in bioinformatics.

The relatively high number of false positive cases detected by the computer-assisted process consisted of truly positive cases that had been missed during manual analysis. The vast majority of discrepancies between computer-assisted and manual classification was for cases classified as false positives by the

computer-assisted process. On reevaluation, these were in fact found to be correctly classified and had been missed by manual evaluation. This finding highlights that a single-operator manual analysis chosen as a gold standard method for free-text analysis is not perfect. Repeating the manual process by the same operator and second operator would have helped to evaluate intra- and inter-operator variability of manual process. Combining the results of 2 operators could have improved the outcome of manual analysis. However, the process of manual analysis was very time consuming, and it was expected that one operator was sufficient for the purpose of the study to compare the computer-assisted and manual processes.

The EMRs used for the study were obtained from veterinary practice. However, despite the anatomical and physiological differences between veterinary and human patients, the terminology used to describe clinical scenarios is very similar if not identical. The slight differences in terminology would be easily addressed during the dictionary definition process, therefore, the method described here would be suitable for use in case-detection research of human patient free-text data.

Limitations of This Process

A limitation of the described method is the component of manual data checking to compile exclusion and reinclusion dictionaries. This requires a considerable effort by the operator, eased by the software, but that remains somewhat proportional to the size of the dataset to be analyzed. When compared with more conventional techniques, using computer-assisted rule-based case definition this time and effort is compensated for by the overall improved performance. A further limitation of this study is the use of a single observer for the manual analysis. Comparing the manual analysis performed independently by 2 operators would have provided a mean of validation of the manual analytic process. Similarly, this study compares manual analysis with the combined computer-assisted process, and the excellent agreement between the 2 methods is acceptable to demonstrate that both methods worked equally well. Finally, since the dictionaries are created from the data, this methodology is suited mostly for retrospective evaluation of EMRs, and if new data is being analyzed, then dictionaries should be updated on the new data to ensure maximal specificity and sensitivity.

A limitation of the validation process described in this study lays in the fact that some patients included in the validation dataset had contributed with a different number of cases. Since this methodology aims at identifying the cases referring to the condition of interest and not the patient affected, ideally the dataset should have included a random selection of rows from the original dataset. However, including a random selection of cases of similar size would have likely resulted in only few or no cases containing text referring to the conditions being investigated. Alternatively, a dataset created from a random selection of cases including an equally large number of cases referring to these conditions would have been too large to be evaluated manually. The decision to include data from a selected number of patients was performed to evaluate this methodology over a wider array of lexical variation.

A further limitation of using WordStat may lie in the cost associated with purchase of the software. In early 2017, the software can be purchased by academics for 695USD (1995USD for governmental organizations and 3795USD for commercial companies), which may at first appear expensive in light of the other software available freely, but the cost may be outweighed by the high sensitivity and specificity offered with this procedure. Whether the methodology of using word frequency list, and inclusion, exclusion, and re-inclusion dictionaries could also adapt to other open source software, should be evaluated in future studies.

Comparison With Prior Work

Excellent specificity and sensitivity was expected as each dictionary was created using a list of words obtained from the dataset and included misspelled and abbreviated terms associated with the category. Considerable effort was required to create a comprehensive inclusion dictionary accounting for all misspellings and abbreviations but was essential to maximize sensitivity of the analytic process. The reinclusion dictionary also further contributed to improving sensitivity where necessary. Although browsing through the word list was a time consuming but pivotal step of the analytic process, it also means that the dictionary is very dataset-specific, and if new data, especially from different clinical practices or veterinarians, is added, then the dictionary should be updated to include new terms. All parameters used to evaluate the performance of the analysis process were superior to most of the current techniques reported in a recent systematic review on case detection from EMRs [2]. The improved performance of the current method is the result of an increased work-load of the operator as the rule-based portion of the analysis to identify negations and context modifying instances is performed somewhat manually. The study by Anholt and colleagues (2014) validated the use of the same software package using rule-based case definition to identify negations and context modifying instances and reported an inferior sensitivity comparable with that of other NLP algorithms [2]. However, that study was aimed at syndromic surveillance where specificity was prioritized over sensitivity in order to minimize false positive rates. This highlights how study design dictates which of these text mining methodologies is more suitable. Future studies aimed at describing disease prevalence or risk factor analysis may have stricter requirements of test performance, and the increased effort of the current method could be justified by the increased reliability of the results produced.

Conclusions

In conclusion, the computer-assisted process is significantly faster once inclusion, exclusion, and reinclusion classification dictionaries are prepared on a dataset of this size while preserving performance at least as good as manual analysis. As all words present in the dataset are used, sensitivity does not appear to be an issue for this method. In terms of optimization of specificity and sensitivity, the use of exclusion and reinclusion dictionaries is useful in situations where there are many false positive and false negative cases. This is achieved simply by evaluating the output search of the inclusion dictionary to identify any significant proportion of erroneous classifications. False positive and negative cases appeared proportionally more common when trying to identify general

syndromes, such as colic, but less common when focusing on specific diagnosis or when looking at drug administration. Future area of research should aim at improving the dictionary definition process to make the process more versatile and adaptable to new data. As technologies improve, this method will probably become obsolete as sensitivities, specificities, PPVs, and NPVs of fully computer-assisted processes, whether rule-based or probabilistic, are likely to improve in the future. These processes should ultimately reduce the operator's effort for dictionary creation and be adaptable to new data.

Acknowledgments

The study was performed with the contribution of John Crawford Endowment Fund of the University of Glasgow and RCVS Knowledge target grant. The crucial role of veterinary practices providing the data is also acknowledged.

Conflicts of Interest

None declared.

References

1. Kao A, Poteet SR, editors. Natural Language Processing and Text Mining. London: Springer; 2007.
2. Ford E, Carroll J, Smith H, Scott D, Cassell J. Extracting information from the text of electronic medical records to improve case detection: a systematic review. J Am Med Inform Assoc 2016 Feb 05;23(5):1007-1015 [FREE Full text] [doi: 10.1093/jamia/ocv180]
3. Publications.parliament. 2013. The dismantled national programme for IT in the NHS URL: https://www.publications.parliament.uk/pa/cm201314/cmselect/cmpubacc/294/294.pdf [accessed 2017-06-05] [WebCite Cache ID 6qzo1f54K]
4. Meystre MS, Savova G, Kipper-Schuler K, Hurdle J. Extracting information from textual documents in the electronic health record: a review of recent research. Yearb Med Inform 2008;3(1):128-144 [FREE Full text]
5. Kushida KC, Nichols D, Jadrnicek R, Miller R, Walsh J, Griffin K. Strategies for de-identification and anonymization of electronic health record data for use in multicenter research studies. Med Care 2012 Jul;50 Suppl:S82-101. [doi: 10.1097/MLR.0b013e3182585355] [Medline: 22692265]
6. Greenhalgh T. Narrative based medicine: narrative based medicine in an evidence based world. BMJ 1999 Jan 30;318(7179):323-325 [FREE Full text] [Medline: 9924065]
7. Salton G. The SMART Retrieval System—Experiments in Automatic Document Processing. Upper Saddle River, NJ: Prentice-Hall, Inc; 1971.
8. Harman HD. The text retrieval conferences (trecs). 1996 Presented at: Fourth Text REtrieval Conference (TREC-4); 1996; Vienna, Virginia p. 373-410 URL: http://dl.acm.org/citation.cfm?id=1119070
9. Heinze HD, Morsch M, Holbrook J. Mining free-text medical records. 2001 Presented at: AMIA Symposium; 2001; Washington, DC p. 254-258.
10. Krallinger M, Valencia A, Hirschman L. Linking genes to literature: text mining, information extraction, and retrieval applications for biology. Genome Biol 2008;9(Suppl 2):S8 [FREE Full text] [doi: 10.1186/gb-2008-9-s2-s8] [Medline: 18834499]
11. Piatetsky-Shapiro G. Knowledge discovery in databases. SIGKDD Explor Newsl 2000 Jan 01;1(2):59-61. [doi: 10.1145/846183.846197]
12. Lam K, Parkin T, Riggs C, Morgan K. Use of free text clinical records in identifying syndromes and analysing health data. Vet Rec 2007;161:547-551. [doi: 10.1136/vr.161.16.547]
13. Anholt RM, Berezowski J, Jamal I, Ribble C, Stephen C. Mining free-text medical records for companion animal enteric syndrome surveillance. Prev Vet Med 2014 Mar 01;113(4):417-422. [doi: 10.1016/j.prevetmed.2014.01.017] [Medline: 24485708]

Abbreviations

EMR: electronic medical record
NPV: negative predictive value
NSAID: non-steroidal anti-inflammatory drug
PPV: positive predictive value

Edited by G Eysenbach; submitted 08.12.16; peer-reviewed by HC Kum, A Benis, A Callahan, A Holzinger; comments to author 13.02.17; revised version received 17.02.17; accepted 08.04.17; published 29.06.17.

Please cite as:
Duz M, Marshall JF, Parkin T
Validation of an Improved Computer-Assisted Technique for Mining Free-Text Electronic Medical Records
JMIR Med Inform 2017;5(2):e17
URL: http://medinform.jmir.org/2017/2/e17/
doi:10.2196/medinform.7123
PMID:28663163

Original Paper

Applying STOPP Guidelines in Primary Care Through Electronic Medical Record Decision Support: Randomized Control Trial Highlighting the Importance of Data Quality

Morgan Price[1,2], MD, PhD, CCFP, FCFP; Iryna Davies[1], MSc; Raymond Rusk[1], PhD; Mary Lesperance[2], PhD; Jens Weber[1,2], PEng, PhD

[1]LEAD Lab, Department of Family Practice, Island Medical Program, University of British Columbia, Victoria, BC, Canada
[2]University of Victoria, Victoria, BC, Canada

Corresponding Author:
Morgan Price, MD, PhD, CCFP, FCFP
LEAD Lab
Department of Family Practice, Island Medical Program
University of British Columbia
Medical Science Building University of Victoria PO Box 1700 STN CSC
Victoria, BC, V8W 2Y2
Canada
Phone: 1 2502167709
Fax: 1 2504725505
Email: morgan@leadlab.ca

Abstract

Background: Potentially Inappropriate Prescriptions (PIPs) are a common cause of morbidity, particularly in the elderly.

Objective: We sought to understand how the Screening Tool of Older People's Prescriptions (STOPP) prescribing criteria, implemented in a routinely used primary care Electronic Medical Record (EMR), could impact PIP rates in community (non-academic) primary care practices.

Methods: We conducted a mixed-method, pragmatic, cluster, randomized control trial in research naïve primary care practices. Phase 1: In the randomized controlled trial, 40 fully automated STOPP rules were implemented as EMR alerts during a 16-week intervention period. The control group did not receive the 40 STOPP rules (but received other alerts). Participants were recruited through the OSCAR EMR user group mailing list and in person at user group meetings. Results were assessed by querying EMR data PIPs. EMR data quality probes were included. Phase 2: physicians were invited to participate in 1-hour semi-structured interviews to discuss the results.

Results: In the EMR, 40 STOPP rules were successfully implemented. Phase 1: A total of 28 physicians from 8 practices were recruited (16 in intervention and 12 in control groups). The calculated PIP rate was 2.6% (138/5308) (control) and 4.11% (768/18,668) (intervention) at baseline. No change in PIPs was observed through the intervention (P=.80). Data quality probes generally showed low use of problem list and medication list. Phase 2: A total of 5 physicians participated. All the participants felt that they were aware of the alerts but commented on workflow and presentation challenges.

Conclusions: The calculated PIP rate was markedly less than the expected rate found in literature (2.6% and 4.0% vs 20% in literature). Data quality probes highlighted issues related to completeness of data in areas of the EMR used for PIP reporting and by the decision support such as problem and medication lists. Users also highlighted areas for better integration of STOPP guidelines with prescribing workflows. Many of the STOPP criteria can be implemented in EMRs using simple logic. However, data quality in EMRs continues to be a challenge and was a limiting step in the effectiveness of the decision support in this study. This is important as decision makers continue to fund implementation and adoption of EMRs with the expectation of the use of advanced tools (such as decision support) without ongoing review of data quality and improvement.

Trial Registration: Clinicaltrials.gov NCT02130895; https://clinicaltrials.gov/ct2/show/NCT02130895 (Archived by WebCite at http://www.webcitation.org/6qyFigSYT)

(JMIR Med Inform 2017;5(2):e15) doi:10.2196/medinform.6226

KEYWORDS

electronic medical records; clinical decision support; randomized control trial; electronic prescribing; data quality

Introduction

Potentially Inappropriate Prescriptions

Adverse Drug Events (ADEs), injuries, and deaths resulting from the administration of a medication [1-3], are a leading cause of iatrogenic morbidity and mortality. Canadian adverse event rates are estimated at 185,000 annually, with 70,000 being potentially preventable [4,5]. The ADE rates are similar in the United States, with the Institute of Medicine estimating that 100,000 preventable deaths occur per year in the United States [6,7]. Medication errors have a greater impact on vulnerable populations such as the elderly, who have significant illness burdens and are often taking a number of medications [8]. It has been reported that 27% of the elderly are on 5 or more medications [8]. The cost of ADEs in seniors is high: over Can $35 million annually in Canada [9]. Avoiding inappropriate prescriptions is one important approach to avoiding predictable ADEs among older people [10]. Effective prevention should involve primary care.

Criteria for the Screening Tool of Older People's Prescriptions

Several groups have attempted to reduce inappropriate prescriptions for the elderly, creating a number of guidelines and criteria to help prescribers use a rational approach to drug prescriptions for the elderly [11]. The Beers criteria [12-16] were developed to support clinicians, and more recently, the STOPP criteria (screening tool of older people's prescriptions) have been developed [17,18]. The STOPP criteria [17-19] consist of 65 recommendations (114 in version 2 [19]) that support evidence-based, individualized prescribing practices among patients 65 and over. The criteria take into account a range of salient patient features to predict potentially inappropriate prescriptions. A systematic review showed that STOPP version 1 was more sensitive than Beers in identifying inappropriate prescribing [11]. The majority of STOPP literature is focused on long-term care and hospital settings [12,14-16,20]. Less work has been done with STOPP in primary care [21,22], even though preventable ADEs are common and serious in this setting [4,23]. In response, this study set out to measure the impact of using the STOPP criteria in primary care, where the majority of prescriptions occur.

Promise of Decision Support

Clinical Decision Support (CDS) aids clinicians and patients in making appropriate decisions in care. In primary care, CDS is often embedded into Electronic Medical Records (EMRs). There is promise for CDS in improving quality of care in general and prescribing in particular, with 66% of studies on prescribing systems showing positive outcomes [24]. Although there is promise in these tools, the benefits of using these tools are not consistently realized [25,26]. Studies, such as the MOXXI study, have shown variable responses to CDS for prescribing [27]. In some cases, the user experience of CDS tools is poor enough that alerts are overridden [1,3], the use of CDS and electronic tools have facilitated errors [4], or the CDS tools have had unintended consequences [28,29]. There is a pressing need to improve decision support tools for providers in order to better realize the expected benefits and reduce serious, unintended consequences.

Achieving impact with CDS tools is not without challenges. There are many "grand challenges" for CDS [30] such as development of content, making content available through Web-based systems, and user experience. The American Medical Informatics Association's position paper on CDS design recommends CDS tools that better summarize and prioritize recommendations to reduce the cognitive burden on clinicians and maintain efficiency [31]. SAFER guidelines have been developed to "empower organizations to work with internal or external stakeholders on optimizing [EMR] functionality" [32]. Data quality is often discussed in terms of data use in research [33]; however, data quality is also foundational to CDS [34].

Research Objective

Through this mixed-method study, we sought to answer the following overarching question:

How can an existing clinical decision support tool implement a complex set of evidence-based rules into primary care clinical practice and how does this impact prescribing?

We considered this question in the context of a primary care EMR with CDS. The EMR was able to be populated by a Web-based decision support application to provide rules. We sought to understand the answers to the question through a combination of EMR data quality probes and participant interviews.

Methods

This was a mixed-method study divided into two phases: (1) A randomized control trial and (2) A qualitative reflection by participants on the results.

Phase 1: Pragmatic Randomized Control Trial

Phase 1 was a prospective, intention-to-treat, un-blinded, cluster randomized trial. The primary outcome measure was the *change* in rate of STOPP-defined PIPs as documented in the OSCAR EMR in the intervention group as compared with the control group.

Participants

Inclusion Criteria

The participants were primary care physicians in British Columbia providing office-based care to patients 65 and over, using the open source OSCAR EMR developed by McMaster University and the OSCAR community (version 12.x) for at least 12 months (this was to provide enough time for medications to be consistently documented in the EMR), and who were part of or willing to be part of the University of British Columbia's Department of Family Practice Research Network.

Exclusion Criteria

Providers who do not provide longitudinal care (eg, walk in clinics) or only hospital care, who do not use OSCAR for writing prescriptions, or who provide care to a younger population (eg, a maternity clinic) were excluded from the study.

Sample Size

Sample size was calculated assuming a PIP rate of 20% [10,27,35] and an expected relative reduction of 20% in PIPs (absolute reduction of 4%). Using a power of 0.8 and alpha=0.05 and estimating that two practices may be lost to follow up, we predicted the need for 12 practices in each arm and 900 encounters per arm with patients 65 years and over.

Recruitment and Randomization

Providers were recruited through the OSCAR Canada User Society's mailing list and the 2014 OSCAR EMR national user meeting (over 100 people were made aware of the study). Potential participants were screened by the primary investigator on the phone or in person to ensure they met the criteria.

Clinics were randomly assigned (equal distribution by clinic) to the intervention or control groups using a random number generator that generated the list prior to the recruitment and allocation of participants. Randomization was stratified into small (<4 physicians) and large clinics (≥4 physicians).

All physicians (in the control and intervention groups) received the same orientation to the purpose of the study, the nature of the STOPP criteria, and what was being measured. The intervention group also received assistance in activating the STOPP criteria in their CDS tools. The control group was invited to have the STOPP guidelines activated after the study and the guidelines were made freely available to the OSCAR community after the study.

Intervention

The intervention group received the STOPP guidelines content in their EMR, whereas the control group did not. The STOPP guidelines leveraged the existing CDS engine of the EMR. These additional guidelines provided suggestions to the providers when specific criteria were met for an individual patient being seen. The EMR showed patient specific guideline recommendation titles in a text window in the side bar of the patient's chart. These titles could be clicked on for more information (Figure 1).

Implementation at the user level was required in the intervention group and facilitated by one of the authors (ID). Participants were instructed and walked through how to turn on the STOPP rules by trusting the STOPP content in the Clinical Decision Support (CDS) module. This downloaded and activated the STOPP rules for that user for the duration of the trial.

A subset of 40 STOPP rules (that were found in both STOPP v1 and v2) was developed for the EMR CDS rules engine. The rules and the network queries (which measured the outcomes as PIP rates) were generated from the same logic files to ensure consistency between the EMR CDS rules and the network measurements.

The 40 STOPP rules were successfully modeled and implemented for this study. It was not possible to create all STOPP rules due to features in the EMR or network query engine. STOPP rules that were not included contained concepts such as duration of combined prescriptions or dose thresholds that could not be modeled in one of the two components (eg, the EMR guideline logic or the network query). Shortly before the study was to start, version 2 of the STOPP rules were published. Rules from version 1 that were removed in version 2 were removed from our study.

Primary Outcome Measure

The primary outcome measure was the difference in change in measured PIP rates between the intervention and control groups before the intervention as compared with the difference after the intervention period.

Data Quality Probes

To provide context and estimate the validity of answers, a set of 13 data quality probes (DQ probes) were created. These assessed the data quality of demographics, medications, and the problem list—the three areas that were in the control of the clinic and related to the STOPP criteria. All DQ probes considered only those patients that had had an encounter at the clinic in the last two years. The DQ probes list is shown in Table 2.

Data Collection

Phase 1: Collection of Physician Information

All the participants completed a survey describing their practice at the start of the study.

Measurement of Potentially Inappropriate Prescriptions

Measurement of PIPs and DQ probes was completed using the UBC Department of Family Practice research network, which was developed with the Physicians Data Collaborative of British Columbia. The network is based on hQuery, an open source tool that is freely available on GitHub. The research network is designed to distribute querying of EMR health data without collection and storage of patient level data. Only aggregate data (ie, summary answers to queries) are collected in the central Hub [36]. Patient privacy was maintained through the network as only practice level aggregate answers were returned through the network.

Baseline Rates of Potentially Inappropriate Prescriptions

Queries were run at the start of the study for each clinic to provide baseline data of the 16 weeks preceding the study.

Intervention Rates of Potentially Inappropriate Prescriptions

Queries were repeated after the 16-week intervention period to assess PIP rates for the intervention period.

Data Analysis

PIPs for each of the STOPP rules and DQ probes were measured for each clinic. Statistical regression models (which account for the clustered nature of the data) were fit to assess the primary outcome (geeglm in R package geepack) [37].

All clinics received an individualized summary of the findings from the study including PIP rates, most common PIPs in their practice, and highlights of data quality in their EMRs.

Phase 2: Explaining Findings and Understanding Experiences

Phase 2 consisted of one-hour semistructured interviews with physicians in the intervention group. These reflection sessions encouraged physicians to explain the findings. The findings for the study were shared, including physicians' own clinic specific summaries, and they were asked to reflect on the results and describe their experiences. Participants were invited to provide information on how to improve the tools in the EMR.

This study was registered at clinical trials.gov (NCT02130895) and received clinical ethics approval from the University of British Columbia (H14-00797).

Results

Participation

The study was completed from February-October, 2015. A total of 8 clinics were engaged in the study (9 were approached, but one declined because of technical reasons with their EMR) and randomization occurred at the clinic level. Twenty-eight physicians across the eight practices consented in person. None were lost to follow up (Figure 2). One participant reported a technical problem with the EMR that was thought to be related to the intervention and later discovered to be unrelated.

Figure 1. Wireframe of the CDS alerts in the EMR. (A) on the right is the panel that lists patient specific alerts. From that panel, users can click a title and get (B), the detail of the alert that pops up when clicked.

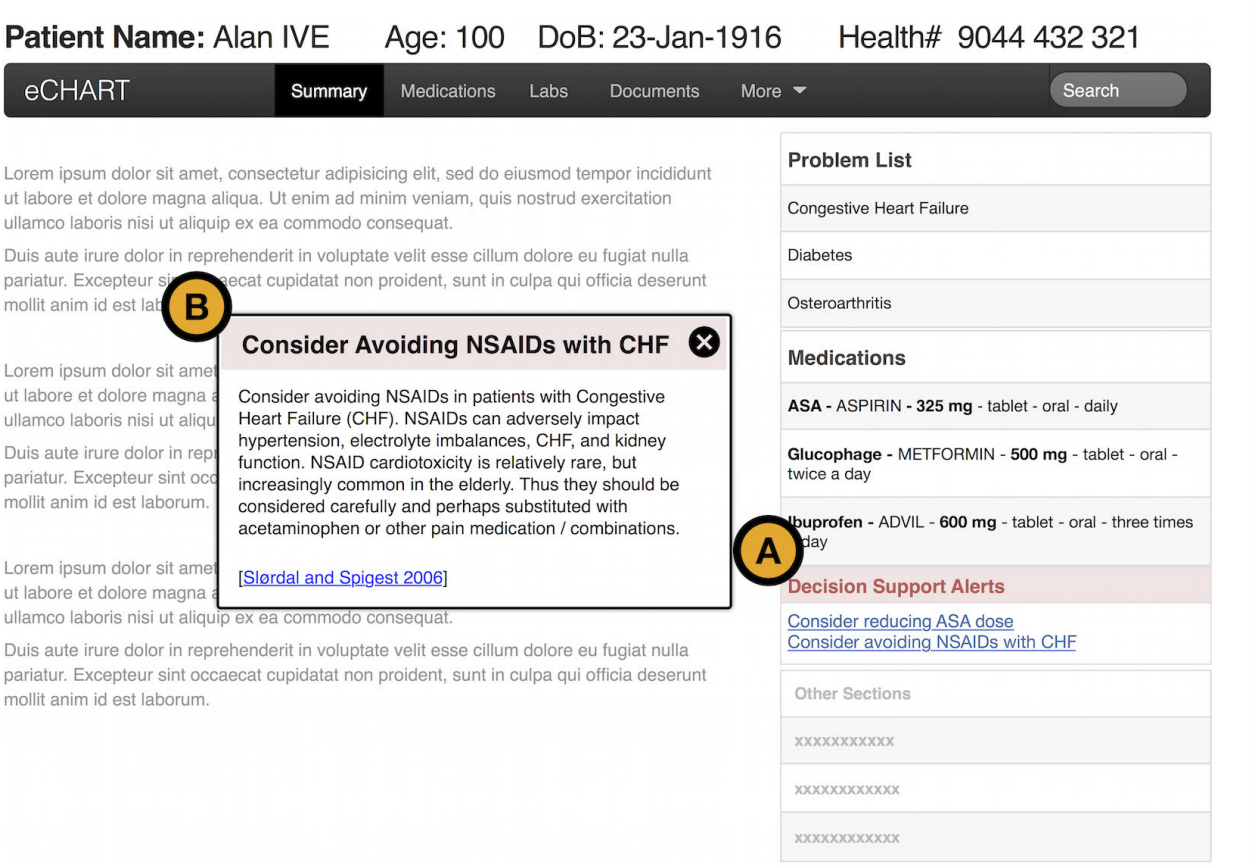
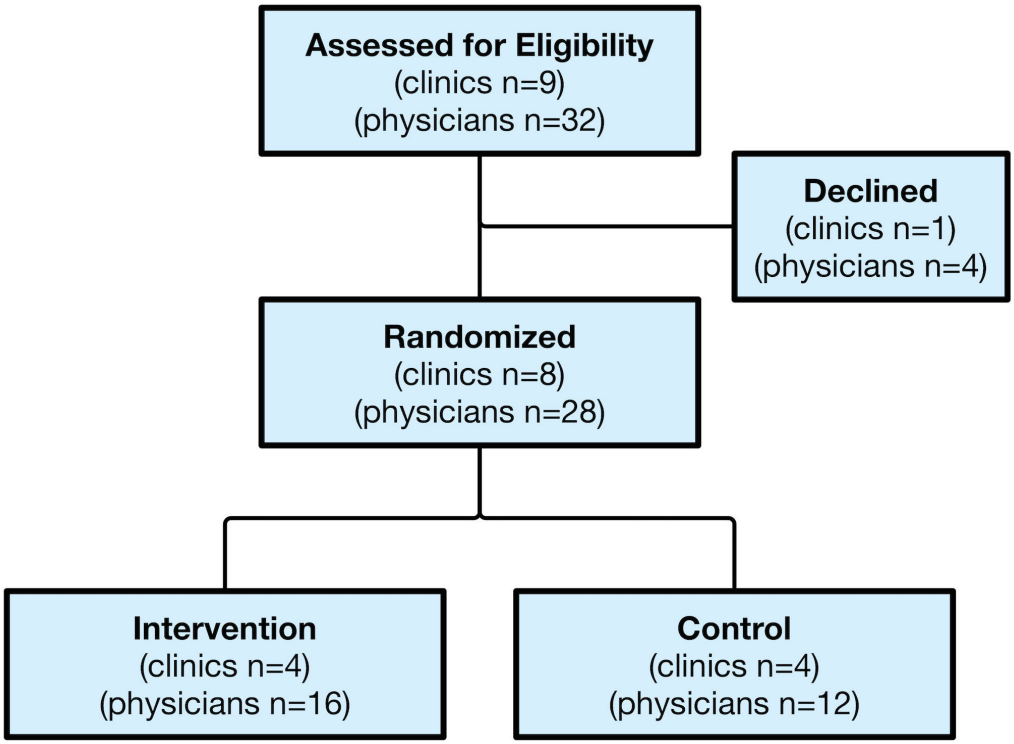


Figure 2. CONSORT figure. 28 physicians in 8 clinics were recruited into the study. 1 clinic declined to participate during recruitment. No clinics or physicians were lost to follow up during the trial.



Phase 1 Results

The control group saw 1086 patients who could have triggered a PIP during the baseline period and 1204 during the treatment period. In the control group, there were 138 PIPs (out of a possible 5308 that could have been triggered) during baseline and 157 PIPs (out of a possible 5792) during treatment.

The intervention group saw 3556 patients who could have triggered a PIP during the baseline period and 3621 during the treatment period. In the intervention group, there were 742 PIPs during baseline (out of a possible 18,331) and 768 PIPs during treatment (out of a possible 18,668). There was an initial difference between the two groups (2.6% and 4.0% of prescriptions were flagged as PIPs in the control and intervention groups, respectively).

Table 1. Rates of PIPs in the control and intervention groups before and during the treatment periods.

Rates	Intervention, %	Control, %
Before treatment (baseline)	4.0	2.6
During treatment	4.1	2.7
Change in potentially inappropriate prescriptions (PIP)rates	0.1	0.1

Both groups saw similar numbers of patients during the study (44,290 in the control group and 37,615 in the intervention group); however, the intervention group had a greater proportion of patients who were 65 years and older (control 5% vs intervention 19%).

The regression analysis of PIPs showed no significant difference in change of recorded PIPs in the control group versus the intervention group (P=.80).

Table 2. Data quality probes assessing demographics, medications, and problem list usage.

Data quality probe description	Intervention, %	Control, %
What percentage of patients, flagged as active, had at least one encounter in the past 24 months?	35.7	28.6
What is the percentage of patients, calculated as active, with no documented gender?	0.3	0.5
What percentage of patients, calculated as active, has an invalid date of birth?	0.0	0.0
What percentage of patients, calculated as active, has no documented date of birth?	0.0	0.0
What percentage of current medications is coded?	82.3	79.0
What percentage of patients, calculated as active, has no current medications?	69.9	84.6
What percentage of problems on the problem list, documented in the past 12 months, has a diagnostic code?	100.0	100.0
What percentage of patients, aged 12 years and over and calculated as active, has at least one documented problem on the problem list (documented in the past 12 months)?	12.3	3.5
What percentage of patients, calculated as active and aged 12 years and over, has Diabetes on the problem list?	5.3	3.5
Of patients currently on Tiotropium medication, what percentage has "COPD"[a] on the problem list?	48.1	72.7
Of patients currently on Levothyroxine medication, what percentage has "Hypothyroidism" on the problem list?	22.7	24.3
Of patients currently on anti-gout medication, what percentage has "Gout" on the problem list?	8.4	12.2

[a]COPD: chronic obstructive pulmonary disease.

Data Quality Probes

EMR data quality was estimated using a set of DQ probes that were executed in the same way as PIP rates were evaluated. The DQ probes were used to assess completeness, correctness, and concordance.

Both groups had a large number of active patients (an average of 68.2%) that were documented as active according to the EMR but that had not had an appointment in at least two years. (Note: STOPP queries were designed only to look at patients who had had an encounter during either the baseline or intervention period). We found that 79-82% of prescriptions were coded, with the remainder being free text; however, there were a high number of patients who were not on any active medications in both groups. All documented problems on the problem list were coded (this was a requirement for the working of the EMR); however, only a small number of patients (12.3% in the intervention group and 3.5% in the control group) had at least one problem on the problem list. In the control group, interestingly, nearly all the patients were also diabetic. The three DQ probes that relate diagnosis to medication use (COPD, hypothyroidism, and gout) all showed that the coded problem list was under-utilized.

The software engineers on this project confirmed the results of the study's queries by using an alternate query method to ensure that the query logic was running correctly. These queries confirmed the above findings.

Phase 2 Findings

Total of 5 physicians across 3 of the 4 intervention group clinics participated in phase 2. Two of the participants had discussed the study and the phase 2 interview with their colleagues in preparation for the meeting and shared their collective thoughts. Three themes emerged from the interviews.

Alert Awareness

All phase 2 participants felt that they were aware of the STOPP alerts (participants were not blinded to being in the intervention group). However, although they all felt that they had seen "some" CDS alerts, they felt they might not have *consistently* seen them.

Workflow and Display

The location on screen and the workflow were thought to be barriers. The STOPP criteria, as they had more complex rules, were implemented differently and displayed in a separate location to simple drug alerts. This often meant that the user would need to tab between screens and refresh screens. Participants preferred a single location for all medication related alerts, regardless of the logic behind those alerts.

Study Disruptiveness

Finally, the participants reflected on the disruption caused by the study to their practice. All of them agreed that the disruption caused by this kind of study was minimal and that they would all participate in future studies.

Discussion

We sought to obtain an answer to the following question:

How can an existing clinical decision support tool implement a complex set of evidence-based rules into primary care clinical practice and how does this impact prescribing?

A total of 40 STOPP criteria were implemented in the primary care EMR. There were some limitations in the CDS module logic that prevented some rules from being implemented, such as being able to calculate the duration for which a patient had been on a medication over multiple prescriptions. However, the randomized trial component of this study was unable to show a significant change in PIPs rates that could be attributed to the

STOPP guidelines as implemented. There are at least two reasons for this.

First, the rate of measured PIPs was lower than expected when compared with PIP rates found in the literature [10,27,35]. With measured PIP rates of 2.6% and 4.0%, it was not possible to see the decrease in PIPs expected from other studies. The DQ probes begin to shed light on why the measured PIP rate is lower than expected. The areas of the EMR that were used by the STOPP rules (eg, the coded problems on the problem list and coded prescriptions that were up to date) were utilized with less frequency than expected and, thus, the CDS rules were not fired as often. For example, we see that 92% of patients had no diagnosis input into their coded problem list, which is a rate of use that is lower than anticipated. Gaps in EMR data quality are as important today as they were twenty years ago [38]. While electronic medical record data are increasingly available and easy to access, the data quality is increasingly the challenge [39]. This is a key point highlighted by this study that should be considered by implementers and decision makers. Even though this study was completed with people using an established EMR that has been widely adopted in Canada and had been adopted in the participating clinics for some time, the data was not fit for use in the CDS.

Data quality is often talked about in terms of data warehouses [33], connecting multiple systems [40], and big data [41]. This study also highlights that data quality is a limiter when applying computational interventions that are designed to support improvements in care for individual patients. Missing, uncoded, or variably documented (eg, in another location within the EMR) data may well have been the limiter in this study.

Second, the participants in phase 2 discussed some challenges with the CDS. For example, the workflow of the STOPP alerts was less than ideal: for technical reasons, the STOPP rules were presented in the main chart and not in the prescription module. The guideline tool did not have a clear way to support users in prioritizing suggestions and alerts as recommended [42]. The alerts were positioned near the bottom of the screen and the users felt that, while they were aware of them, they were small, in a list with other reminders, and could potentially be missed.

Despite the negative result, there are still several valuable lessons that can be learned from this pragmatic trial.

Lessons Learned

Translating Rules for the Screening Tool of Older People's Prescriptions into Electronic Medical Record Algorithms

In general, the STOPP rules were well specified and computable definitions that could be created for most rules. However, in many cases, definitions had to be refined based on the logic features and data accessible in the EMR. The STOPP rules were not defined using specific medical terminologies (eg, International Classification of Disease (ICD) codes) and these had to be developed for this study. This gap is not unique to STOPP rules; indeed, others have had similar challenges in translating clinical guidelines into computer interpretable rules or alerts [43]. However, we recommend that future versions of STOPP and other recommendations consider using terminologies in their definitions to aid in consistent translation into computable forms. Finally, the CDS engine was not able to access all EMR data elements described in the STOPP criteria and several rules were excluded.

Data Quality

This was a pragmatic trial. We wanted to see the impact of implementation of CDS in nonacademic practices that had not gone through extensive data quality improvement training. This would better predict impact of CDS tools in real world settings. We discovered that 77% of patients were not on medications that could be queried. Approximately 20% of prescriptions were documented in the prescription writer without codes that could be queried, indicating custom medications (ie, free text) that would further limit the ability of the CDS. Over 92% of patients had no coded problems on the problem list (these were queries of the practice, not just the study population). As these were two data sources for many of the queries, data quality was one of the limiting factors to the triggering of the CDS guidelines. This would likely impact other similar uses of CDS without remediation. The EMR allowed for free text in several places (eg, within encounter notes) and so it would be possible for many more items to be recorded in the EMR than could be queried in a coded manner, if the respective EMR components were not used as intended.

Participants in phase 2 were generally unaware of these data quality issues. Previous work has highlighted that physicians often think of the EMR as an "electronic paper record" [44] and do not consider the downstream impacts of not using or inappropriately using components of the EMR. There is an ongoing need for more general education around the use of EMR and EMR data for primary care physicians.

EMR Workflow Limitations

In this study, the EMR could not integrate complex CDS rules into the prescribing module. Thus, STOPP alerts appeared in the main screen of the patient chart. Future implementations should allow for better integration into key workflows, such as prescribing. Furthermore, while the guideline engine did allow for summary and additional information to be displayed on user request, it did not allow for easy actions to be performed (eg, discontinuing a medication that triggered the STOPP guideline). There was an EMR workflow that users could use that would skip the screen where the CDS STOPP rules were displayed; however, participants in phase 2 stated that this was not a typical workflow. This issue has been addressed in an upcoming version of the EMR.

Study Implementation

One of the guiding principles when developing this study was to implement and run the study with as little disruptive impact on practicing physicians as possible. Feedback from the participants was positive in terms of ease of study participation. However, an important matter to explore in future work is the minimum amount of training needed to achieve sufficient improvements in data quality to achieve benefits from the application of complex rules like STOPP. This has implications not just for future studies but also for the implementation of

future CDS as part of EMR requirements and quality improvement initiatives.

EMRs Permit Variable Workflows

Although we confirmed subjectively through phase 2 interviews that practitioners saw the guidelines and had the opportunity to act, we did not have a mechanism to proactively measure individual workflows in the EMR. It is possible (eg, through quick links for direct medication renewals) to avoid the triggering of the EMR guideline module where the STOPP rules reside. We were not able to measure how often this quick link or other paths might have been used. This information would be helpful in understanding how to redesign the EMR and other clinical information systems in the future.

Embedding Knowledge Translation

This study was designed with two knowledge translation (KT) partners: the EMR and the research network. The STOPP alerts and network queries were developed with these two groups and the materials were provided to each group freely at the end of the study. This proved to be an effective way of engaging with participants and partners and ensuring that the knowledge and artifacts from the study have future application. It was an excellent model to engage both the partners and the end-users in the study as the providers understood that their participation would allow for greater and ongoing impacts in the community.

Study Limitations

The study had several limitations, which were as follows. (1) Participants were not blinded to which arm of the study they were in. (2) Physicians and clinics volunteered to participate in the network and the study. (3) Because of the timing of the study, the newest version of the particular EMR was not yet installed for the study clinics. This resulted in using an outdated user interface model. The newer EMR has addressed several workflow issues related to CDS and prescribing. (4) Only 40 STOPP study criteria were implemented. (5) Data quality probes were of a more general nature and not specific to the age range of this study's patient population.

Future Direction

Data quality is a key issue for the use of CDS tools, especially outside of large academic centers where there may be additional resources to improve data quality. The authors have begun to consider data quality by design as an engineering framework that can be applied in the real world [45]. Given the highlights related to workflow and CDS from this study, current work is exploring new user interface designs that support different paradigms for CDS [46] in prescribing. These design ideas are being shared with EMR vendors. As newer versions of the EMR and other study components are engineered and adopted, repeating the study will allow for some level of comparison to assess the changes in design.

Conclusions

This pragmatic study intentionally implemented a subset of STOPP prescribing guidelines into nonacademic primary care offices with minimal training and disruption. One of the limitations discovered was the data quality in the EMR databases. The rates of measured potentially inappropriate prescriptions was limited and, thus, the rate at which the decision support would be triggered was limited by insufficient use of the EMR components that were connected to the decision support system. Further, this study provides more evidence to support the need to carefully design the workflows of the EMR tools that will support quality. As decision makers create policy to implement tools in EMRs such as decision support, careful attention will be required to ensure that practices and their data are ready to adopt these tools.

Acknowledgments

We would like to thank the Canadian Institutes of Health Research that funded this research through the eHealth Catalyst Grant program; the Physicians Data Collaborative of British Columbia; OSCAR-EMR and the OSCAR Canada User Society, who engaged and supported this work; and the UBC Department of Family Practice's practice-based research network.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT EHealth Checklist (V 1.6.1).

[PDF File (Adobe PDF File), 730KB - medinform_v5i2e15_app1.pdf]

References

1. Weingart SN, Toth M, Sands DZ, Aronson MD, Davis RB, Phillips RS. Physicians' decisions to override computerized drug alerts in primary care. Arch Intern Med 2003 Nov 24;163(21):2625-2631. [doi: 10.1001/archinte.163.21.2625] [Medline: 14638563]
2. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. J Am Med Assoc 1998 Apr 15;279(15):1200-1205. [Medline: 9555760]
3. Shah SGS, Robinson I. Benefits of and barriers to involving users in medical device technology development and evaluation. Int J Technol Assess Health Care 2007;23(1):131-137. [doi: 10.1017/S0266462307051677] [Medline: 17234027]

4. Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, et al. Role of computerized physician order entry systems in facilitating medication errors. J Am Med Assoc 2005 Mar 9;293(10):1197-1203. [doi: 10.1001/jama.293.10.1197] [Medline: 15755942]
5. Baker GR, Norton PG, Flintoft V, Blais R, Brown A, Cox J, et al. The Canadian Adverse Events study: the incidence of adverse events among hospital patients in Canada. CMAJ 2004 May 25;170(11):1678-1686 [FREE Full text] [Medline: 15159366]
6. Richardson WC, Berwick DM, Bisgard JC, Bristow LR, Buck CR, Cassel CK. Crossing the quality chasm: a new health system for the 21st century. Washington, D.C: National Academy Press; 2001.
7. Budnitz DS, Lovegrove MC, Shehab N, Richards CL. Emergency hospitalizations for adverse drug events in older Americans. N Engl J Med 2011 Nov 24;365(21):2002-2012. [doi: 10.1056/NEJMsa1103053] [Medline: 22111719]
8. Reason B, Terner M, Moses MA, Tipper B, Webster G. The impact of polypharmacy on the health of Canadian seniors. Fam Pract 2012 Aug;29(4):427-432. [doi: 10.1093/fampra/cmr124] [Medline: 22223743]
9. Wu C, Bell CM, Wodchis WP. Incidence and economic burden of adverse drug reactions among elderly patients in Ontario emergency departments: a retrospective study. Drug Saf 2012 Sep 01;35(9):769-781 [FREE Full text] [doi: 10.2165/11599540-000000000-00000] [Medline: 22823502]
10. Ryan C, O'Mahony D, Kennedy J, Weedle P, Byrne S. Potentially inappropriate prescribing in an Irish elderly population in primary care. Br J Clin Pharmacol 2009 Dec;68(6):936-947 [FREE Full text] [doi: 10.1111/j.1365-2125.2009.03531.x] [Medline: 20002089]
11. Hill-Taylor B, Sketris I, Hayden J, Byrne S, O'Sullivan D, Christie R. Application of the STOPP/START criteria: a systematic review of the prevalence of potentially inappropriate prescribing in older adults, and evidence of clinical, humanistic and economic impact. J Clin Pharm Ther 2013 Oct;38(5):360-372. [doi: 10.1111/jcpt.12059] [Medline: 23550814]
12. Pyszka LL, Seys RT, Milhans SM. Identification of inappropriate prescribing in geriatrics at a Veterans Affairs hospital using STOPP/START screening tools. Consult Pharm 2010 Jun;25(6):365-373. [doi: 10.4140/TCP.n.2010.365] [Medline: 20534407]
13. By the American Geriatrics Society 2015 Beers Criteria Update Expert Panel. American Geriatrics Society 2015 updated beers criteria for potentially inappropriate medication use in older adults. J Am Geriatr Soc 2015 Nov;63(11):2227-2246. [doi: 10.1111/jgs.13702] [Medline: 26446832]
14. García-Gollarte F, Baleriola-Júlvez J, Ferrero-López I, Cruz-Jentoft AJ. Inappropriate drug prescription at nursing home admission. J Am Med Dir Assoc 2012 Jan;13(1):83.e9-83.15. [doi: 10.1016/j.jamda.2011.02.009] [Medline: 22208763]
15. Liu C, Peng L, Chen Y, Lin M, Liu L, Chen L. Potentially inappropriate prescribing (IP) for elderly medical inpatients in Taiwan: a hospital-based study. Arch Gerontol Geriatr 2012;55(1):148-151. [doi: 10.1016/j.archger.2011.07.001] [Medline: 21820189]
16. Parsons C, Johnston S, Mathie E, Baron N, Machen I, Amador S, et al. Potentially inappropriate prescribing in older people with dementia in care homes: a retrospective analysis. Drugs Aging 2012 Feb 01;29(2):143-155. [doi: 10.2165/11598560-000000000-00000] [Medline: 22204669]
17. Gallagher P, O'Mahony D. STOPP (Screening Tool of Older Persons' potentially inappropriate Prescriptions): application to acutely ill elderly patients and comparison with Beers' criteria. Age Ageing 2008 Nov;37(6):673-679 [FREE Full text] [doi: 10.1093/ageing/afn197] [Medline: 18829684]
18. Gallagher P, Ryan C, Byrne S, Kennedy J, O'Mahony D. STOPP (Screening Tool of Older Person's Prescriptions) and START (Screening Tool to Alert doctors to Right Treatment). Consensus validation. Int J Clin Pharmacol Ther 2008 Feb;46(2):72-83. [Medline: 18218287]
19. O'Mahony D, O'Sullivan D, Byrne S, O'Connor MN, Ryan C, Gallagher P. STOPP/START criteria for potentially inappropriate prescribing in older people: version 2. Age Ageing 2015 Mar;44(2):213-218 [FREE Full text] [doi: 10.1093/ageing/afu145] [Medline: 25324330]
20. Hamilton HJ, Gallagher PF, O'Mahony D. Inappropriate prescribing and adverse drug events in older people. BMC Geriatr 2009 Jan 28;9:5 [FREE Full text] [doi: 10.1186/1471-2318-9-5] [Medline: 19175914]
21. Yayla ME, Bilge U, Binen E, Keskin A. The use of START/STOPP criteria for elderly patients in primary care. ScientificWorldJournal 2013;2013:165873 [FREE Full text] [doi: 10.1155/2013/165873] [Medline: 23853529]
22. Conejos MM, Sánchez CM, Delgado SE, Sevilla MI, González-Blazquez S, Montero EB. Potentially inappropriate drug prescription in older subjects across health care settings. Eur Geriatr Med 2010 Feb 01;1(1):9-14. [doi: 10.1016/j.eurger.2009.12.002]
23. Gurwitz JH, Field TS, Harrold LR, Rothschild J, Debellis K, Seger AC, et al. Incidence and preventability of adverse drug events among older persons in the ambulatory setting. J Am Med Assoc 2003 Mar 5;289(9):1107-1116. [Medline: 12622580]
24. Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. J Am Med Assoc 2005 Mar 9;293(10):1223-1238. [doi: 10.1001/jama.293.10.1223] [Medline: 15755945]
25. Lau F, Kuziemsky C, Price M, Gardner J. A review on systematic reviews of health information system studies. J Am Med Inform Assoc 2010;17(6):637-645 [FREE Full text] [doi: 10.1136/jamia.2010.004838] [Medline: 20962125]

26. Lau F, Price M, Boyd J, Partridge C, Bell H, Raworth R. Impact of electronic medical record on physician practice in office settings: a systematic review. BMC Med Inform Decis Mak 2012 Feb 24;12:10 [FREE Full text] [doi: 10.1186/1472-6947-12-10] [Medline: 22364529]
27. Tamblyn R, Huang A, Taylor L, Kawasumi Y, Bartlett G, Grad R, et al. A randomized trial of the effectiveness of on-demand versus computer-triggered drug decision support in primary care. J Am Med Inform Assoc 2008;15(4):430-438 [FREE Full text] [doi: 10.1197/jamia.M2606] [Medline: 18436904]
28. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc 2004 Apr;11(2):104-112 [FREE Full text] [doi: 10.1197/jamia.M1471] [Medline: 14633936]
29. Yu P, Zhang Y, Gong Y, Zhang J. Unintended adverse consequences of introducing electronic health records in residential aged care homes. Int J Med Inform 2013 Sep;82(9):772-788. [doi: 10.1016/j.ijmedinf.2013.05.008] [Medline: 23770027]
30. Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, et al. Grand challenges in clinical decision support. J Biomed Inform 2008 Apr;41(2):387-392 [FREE Full text] [doi: 10.1016/j.jbi.2007.09.003] [Medline: 18029232]
31. Middleton B, Bloomrosen M, Dente MA, Hashmat B, Koppel R, Overhage JM, American Medical Informatics Association. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. J Am Med Inform Assoc 2013 Jun;20(e1):e2-e8 [FREE Full text] [doi: 10.1136/amiajnl-2012-001458] [Medline: 23355463]
32. Sittig DF, Ash JS, Singh H. The SAFER guides: empowering organizations to improve the safety and effectiveness of electronic health records. Am J Manag Care 2014 May;20(5):418-423 [FREE Full text] [Medline: 25181570]
33. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc 2013 Jan 1;20(1):144-151 [FREE Full text] [doi: 10.1136/amiajnl-2011-000681] [Medline: 22733976]
34. McCormack JL, Ash JS. Clinician perspectives on the quality of patient data used for clinical decision support: a qualitative study. AMIA Annu Symp Proc 2012;2012:1302-1309 [FREE Full text] [Medline: 23304409]
35. Giguere A, Labrecque M, Grad R, Cauchon M, Greenway M, Légaré F, et al. Barriers and facilitators to implementing Decision Boxes in primary healthcare teams to facilitate shared decisionmaking: a study protocol. BMC Med Inform Decis Mak 2012 Aug 06;12:85 [FREE Full text] [doi: 10.1186/1472-6947-12-85] [Medline: 22867107]
36. Price M, Weber J, McCallum G. SCOOP - The Social Collaboratory for Outcome Oriented Primary Care. : IEEE; 2014 Presented at: IEEE International Conference on Healthcare Informatics (ICHI); 2014; Verona p. 210. [doi: 10.1109/ICHI.2014.36]
37. Halekoh U, Højsgaard S, Yan J. The R package geepack for generalized estimating equations. J Stat Soft 2006;15(2). [doi: 10.18637/jss.v015.i02]
38. Wagner MM, Hogan WR. The accuracy of medication data in an outpatient electronic medical record. J Am Med Inform Assoc 1996;3(3):234-244 [FREE Full text] [Medline: 8723614]
39. Cohen B, Vawdrey DK, Liu J, Caplan D, Furuya EY, Mis FW, et al. Challenges associated with using large data sets for quality assessment and research in clinical settings. Policy Polit Nurs Pract 2015 Aug;16(3-4):117-124 [FREE Full text] [doi: 10.1177/1527154415603358] [Medline: 26351216]
40. Walker KL, Kirillova O, Gillespie SE, Hsiao D, Pishchalenko V, Pai AK, et al. Using the CER Hub to ensure data quality in a multi-institution smoking cessation study. J Am Med Inform Assoc 2014;21(6):1129-1135 [FREE Full text] [doi: 10.1136/amiajnl-2013-002629] [Medline: 24993545]
41. Sukumar SR, Natarajan R, Ferrell RK. Quality of Big Data in health care. Int J Health Care Qual Assur 2015;28(6):621-634. [doi: 10.1108/IJHCQA-07-2014-0080] [Medline: 26156435]
42. Price M, Lau F. The clinical adoption meta-model: a temporal meta-model describing the clinical adoption of health information systems. BMC Med Inform Decis Mak 2014 May 29;14:43 [FREE Full text] [doi: 10.1186/1472-6947-14-43] [Medline: 24884588]
43. Berner ES, Kasiraman RK, Yu F, Ray MN, Houston TK. Data quality in the outpatient setting: impact on clinical decision support systems. AMIA Annu Symp Proc 2005:41-45 [FREE Full text] [Medline: 16778998]
44. Price M, Singer A, Kim J. Adopting electronic medical records: are they just electronic paper records? Can Fam Physician 2013 Jul;59(7):e322-e329 [FREE Full text] [Medline: 23851560]
45. Weber JH, Price M, Davies I. Taming the data quality dragon--a theory and method for data quality by design. Stud Health Technol Inform 2015;216:928. [Medline: 26262230]
46. Price M, Weber JH, Davies I, Bellwood P. Lead user design: medication management in electronic medical records. Stud Health Technol Inform 2015;216:237-241. [Medline: 26262046]

Abbreviations

CDS: clinical decision support
CDSS: clinical decision support system
COPD: Chronic Obstructive Pulmonary Disease

DQ: data quality
ICD: International Classification of Disease
EMR: electronic medical record
PIP: potentially inappropriate prescriptions
RCT: randomized controlled trial
STOPP: screening tool of older people's prescriptions

Edited by G Eysenbach; submitted 17.06.16; peer-reviewed by T Liaw, C Fincham, M Greiver; comments to author 17.11.16; revised version received 21.03.17; accepted 28.04.17; published 15.06.17.

Please cite as:
Price M, Davies I, Rusk R, Lesperance M, Weber J
Applying STOPP Guidelines in Primary Care Through Electronic Medical Record Decision Support: Randomized Control Trial Highlighting the Importance of Data Quality
JMIR Med Inform 2017;5(2):e15
URL: http://medinform.jmir.org/2017/2/e15/
doi:10.2196/medinform.6226
PMID:28619704

Original Paper

Design, Implementation, and Evaluation of Self-Describing Diabetes Medical Records: A Pilot Study

Azadeh Kamel Ghalibaf[1], MSc; Zahra Mazloum Khorasani[2], MD; Mahdi Gholian-Aval[3], MD, PhD; Hamed Tabesh[1], PhD; Mahmood Tara[1], MD, PhD

[1]School of Medicine, Department of Medical Informatics, Mashhad University of Medical Sciences, Mashhad, Islamic Republic Of Iran

[2]School of Medicine, Endocrine Research Center, Mashhad University of Medical Sciences, Mashhad, Islamic Republic Of Iran

[3]School of Health, Department of Health Education and Health Promotion, Mashhad University of Medical Sciences, Mashhad, Islamic Republic Of Iran

Corresponding Author:
Mahmood Tara, MD, PhD
School of Medicine
Department of Medical Informatics
Mashhad University of Medical Sciences
Pardis Daneshgah
Park Square
Mashhad, 917-7948-564
Islamic Republic Of Iran
Phone: 98 5138002429
Fax: 98 5138002445
Email: TaraM@mums.ac.ir

Abstract

Background: Each patient's medical record consists of data specific to that patient and is therefore an appropriate source to adapt educational information content.

Objectives: This study aimed to design and implement an information provision system based on the medical records of diabetic patients and to investigate the attitudes of users toward using this product.

Methods: The study was organized into three phases: need analysis, design and implementation, and final evaluation. The aim of the need analysis phase was to investigate the questioning behavior of the patient in the real-world context. The design and implementation phase consisted of four stages: determining the minimum dataset for diabetes medical records, collecting and validating content, designing and implementing a diabetes electronic medical record system, and data entry. Evaluating the final system was done based on the constructs of the technology acceptance model in the two dimensions of perceived usefulness and perceived ease of use. A semistructured interview was used for this purpose.

Results: Three main categories were extracted for the patient's perceived usefulness of the system: raising the self-awareness and knowledge of patients, improving their self-care, and improving doctor-patient interaction. Both patients and physicians perceived the personalized sense of information as a unique feature of the application and believed that this feature could have a positive effect on the patient's motivation for learning and using information in practice. Specialists believed that providing personal feedback on the patient's lab test results along with general explanations encourages the patients to read the content more precisely. Moreover, accessing medical records and helpful notes was a new and useful experience for the patients.

Conclusions: One of the key perceived benefits of providing tailored information in the context of medical records was raising patient awareness and knowledge. The results obtained from field observations and interviews have shown that patients were ready to accept the system and had a positive attitude when it was put into practice. The findings related to user attitude can be used as a guideline to design the next phase of the research (ie, investigation of system effectiveness on patient outcomes).

(JMIR Med Inform 2017;5(2):e10) doi:10.2196/medinform.6862

KEYWORDS

diabetes electronic medical record; information tailoring; patient education; technology acceptance model

Introduction

Diabetes is a chronic metabolic disease that is highly prevalent in Iran as well as the entire world. Type-2 diabetes is among the most common chronic illnesses in Iran, with a prevalence rate greater than 14% for the population aged 30 years and older [1]. There is no cure for this disease and it requires continuous lifelong care. Results from previous research have shown that raising the knowledge of diabetic patients and their awareness empowers them to better manage their disease [2,3].

Medical records are rich sources for the medical background of patients, and thus can be used as a basis for making medical decisions. In recent years, with the advent of electronic medical records (EMR), the issue of a patient accessing one's medical record has gained more attention than ever before. According to the Health Insurance Portability and Accountability Act of 1996 (HIPAA), patients must be able to see a copy of their records [4]. The Institute of Medicine also advocates for unrestricted patient access to medical records. Existing literature suggests that providing patients access to their medical records can improve their knowledge, facilitate a more collaborative relationship between the provider and patient, patient adherence, greater patient involvement in self-care, and more satisfaction [5,6]. The key point is that the realization of the above-mentioned advantages concerning patient-accessible medical records is possible only when the record content is comprehensible to the patient. A relevant review article [5] has shown that patients commonly had difficulty understanding at least part of their records. Providing explanations about the record not only helps to improve the self-awareness of one's disease status, but also enjoys an educational aspect to help promote patient knowledge and skills, especially if the information provided is structured and relevant to each patient's medical conditions and needs. According to the elaboration likelihood model, when people perceive relevant information about their conditions, they become further motivated to read it and remember more details [7].

Information tailoring systems use an internal representation of user conditions and needs, which is referred to as a "user model" or "user profile." A user model represents the system's beliefs about the user. Hence, it may simply contain demographic information or sophisticated factors such as the state of the disease, user's attitude, interest, preference, and knowledge [8]. Decisions on the number and types of factors involved in the user model depend on the desired dosage of tailoring and the purpose of the system.

Tailoring research began appearing in the literature in the 1990s, and with advances in computer technology, has increased dramatically in recent years [9]. On the basis of the researchers' primary discipline and expertise, we have found two main approaches in this field. The approach led by the computer science community is focused on advanced and intelligent technological methods, but deals less with real-world issues such as implementation context, target user characteristics, and outcome evaluation. Such projects as LEAF (Layman Education and Activation Form) [10], HealthDoc [11], and Migraine [12] are examples of tailored document generation systems within this category.

The second approach relies greatly on evaluating the efficiency of tailored materials compared with generic ones, but generally uses simpler technological methods to generate automatically the tailored print-based materials. A common drawback of research done with this approach is insufficient reporting of tailoring system architecture, which is referred to as the "black box" [9]. We have attempted to fill the existing gap between these two approaches by describing the design and implementation of the system in more detail.

We used the medical record data of patients as the tailoring profile, in which the relevant information retrieved from the library of content is provided to the patient. The system was designed as stand-alone software that can be used in almost any health care environment regardless of the availability of the patient's electronic record. In the presence of an electronic record, our system can be used as an add-on module to the existing EMR, automatically retrieving the patient's data from the EMR. However, there might be a need for simple middleware to convert and adapt the recorded items in EMR to those defined in the user's profile. Embedding the system in the existing EMR system can facilitate its acceptance and practicality.

Research has shown that the target audience's acceptance of the system is a key factor in the success of an information system within a clinical setting. There are two groups of target users whose benefits should be taken into account: final users, who work directly with the system, and clinicians, who confirm and validate the system and support its usage in practice. User experience is a field of study that analyzes the user's emotions, behaviors, and attitudes toward using a product, system, or service [13].

In recent years, different models and frameworks have been introduced to evaluate information systems [14,15]. Technology acceptance model (TAM) is considered the most influential and commonly employed theory for describing an individual's acceptance of information systems. It has been adapted from the theory of reasoned action, which is based on the assumption that one's intention to act out a certain behavior, such as using the system or reading the provided information, is predictable from one's attitude toward that behavior [16]. TAM is the basis for evaluating our system.

Our aim was to design and implement a tailored information provision system based on diabetic patient medical records and to investigate user attitudes toward this product. The research plan was conducted in three phases: need analysis, design and implementation, and final evaluation of the system. First, we have described the methodology used in each phase of the study. Then we have reported the findings of each phase, and finally, we discussed the results and experiences gained along with the causes and factors involved.

Methods

In this section, the methodology used in each phase is explained respectively.

Phase I: Need Analysis

We used field observations of patients in order to identify and prioritize the information needs of the target population. Observation was done in two distinct settings: physicians' visiting sessions for 22 hours, and diabetes educational classes for 10 hours. In the first setting, the researcher took a nonparticipant observer's role in doctor-patient visits and only took field notes. In the second setting, the researcher adopted a patient's role and participated in classes like other ordinary participants. She took notes of the patients' questions and investigated intragroup behavior. We followed a two-step approach to analyze the noted data. We first assigned each question a label to indicate the main topic or concern, and then classified questions based on their labels. In the second step, we performed content analysis for each question to identify the underlying concerns. The findings have been reported in the results section.

Phase II: Design and Implementation

The designing procedures of the tailored self-describing diabetes medical records were conducted in four stages as described in detail in this section.

Stage I: Determining the Minimum Dataset for Diabetes Medical Records

In order to identify a set of items usually recorded in the visits of diabetic patients, the sample paper-based records from five diabetes clinics in Mashhad city were gathered and analyzed. After *eliminating duplicated items*, a checklist was designed with five main categories: demographic information, symptoms, medical history, medication, and lab tests and measurements, with a number of subcategories for each. The checklist was provided to three subspecialists in endocrinology to determine the necessity of including each item. They could also add new items into the "others" section if needed. Items selected by the majority of specialists in the first round of Delphi, along with the newly added items, comprised a checklist for the second round of Delphi. Finally, 110 items were confirmed by the expert panel, including 10 items in the demographic category, 25 items in the symptoms category, 15 items in the medical history category, 45 items in the medication category, and 15 items in the lab test category, all of which included the minimum dataset for diabetes medical records.

Stage II: Content Collection and Validation

We initially needed a structure for collecting and presenting content that corresponded to items in each category. These structures were obtained based on the results of the need analysis phase. As an example, the structure of information for the items in the symptoms category includes four sections: symptom's definition, diagnostic symptoms, causes of occurrence, and advice on prevention and treatment. The information content was collected from credible online and printed sources. Once the comprehensibility and simplification rules—such as lexical simplification, sentence shortening, bullet-points, and so on—were applied to the corpus of the text, we had to validate its quality and correctness. To do this, we provided a printed version of explanations based on the patient's record with a quality assessment checklist to a small representative sample of clinicians and patients. Participants could also comment in free texts. Once the clinicians confirmed the information, it was stored in the system's database.

Stage III: Design and Implementation of Diabetes EMR

The system works in two modes: doctor and patient. We used Microsoft Visual Studio and Microsoft Access database for development.

During each visit, the doctor-user should facilitate the data entry process by entering the patient's data into the system by choosing from among the existing options in the program interface. The doctor can also observe the information about the patient's previous visits by selecting the visit date from the presented list on the screen.

The patient-user can also log in to the system with his or her username and see the record content in two ways (screenshots are available in Multimedia Appendix 1): session-based and topic-based.

Session-based content shows a complete list of all visits based on the visit date. A click on each date reveals all data values recorded on that particular session in three distinct columns titled symptoms and diseases, medication, and lab tests. Each record item is in the form of hypertext; it is clickable and capable of providing explanations. *Topic-based* content provides a visual abstract overview of the chronological progress for the selected data elements in a patient's record.

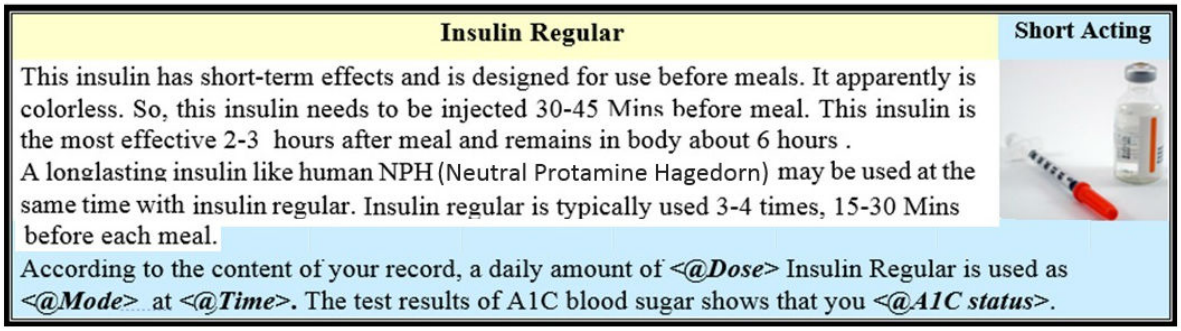
In our proposed EMR system there is a possibility of providing explanations for each item in the patient's record. Clicking on a term opens up a window with explanations. The content of the explanation can be divided into two sections based on the degree of tailoring.

The first section is less tailored, which includes relevant general knowledge about the selected item, and is presented uniformly to all patients. Technically speaking, canned text is used to produce the content of this section, which is stored in advance in the system's library of content. In case all conditions are met, this text is retrieved.

The second section is highly tailored, which interprets the patient's medical condition based on the content of his or her record. To generate the text in this section, the shallow natural language generation approach is used with schema or fill-in-the-blanks strategy.

For example, once the patient clicks on "Insulin Regular," at first a general explanation is provided about this type of insulin and its mechanism of action in the body. Then the second part is focused on a tailored explanation about how "Insulin Regular" should be administered by that specific patient in terms of dosage, timing, and so on (Figure 1; this page originally contained Persian text, shown here is a translation). This type of information provision is based on the assumption that integrating personal information from the patient's medical record with general medical knowledge facilitates a better learning and higher self-awareness of the patient's health status.

Figure 1. System explanations on insulin regular.



Stage IV: Data Injection Into the System

As there were no EMR in the clinic where we conducted the study, we had to manually enter data from patients' paper-based records. To do this we used the clinic appointment-scheduling list to identify patients who would visit in the upcoming days. We then retrieved their medical records from the archive and manually entered the data into the system. On the basis of each patient's record, a personalized document was produced by the system comprising the relevant information based on that patient's medical status. The document was then printed and delivered to the patient in the form of a booklet in his or her upcoming visit session.

Phase III: Evaluation

This phase deals with investigating the users' attitude and perceived value of using the system in practice. The study was conducted in one of the leading health centers providing diabetes prevention, treatment, and management services in Mashhad city. The center was undergoing a switching transition from the traditional paper-based workflow to a digital format at the time of this study.

The evaluation was conducted with two groups of users: doctors and patients. We aimed to identify the key factors involved in the acceptance of the system by users. The methodology used was a semistructured interview based on the constructs of TAM. This model claims that one's acceptance of an information system and intention to use it depends on the two constructs: perceived usefulness and perceived ease of use [17].

Perceived usefulness means one's belief that using the system improves his or her performance. To assess this construct, interview questions were specified in advance in the interview guide based on Davis's standard questionnaire that measures the scales defined in TAM [18] (a detailed description is provided in Multimedia Appendix 2). Interviewees were free to state their opinions about multiple aspects. During each interview session, the interviewer could ask more detailed and precise questions based on the respondent's feedback. Sample questions are: "In your opinion, what advantages are involved in using this system?" "How does the system facilitate your needs?" or "In what ways can the use of this system be useful and helpful?"

Perceived ease of use refers to one's expectation of the system's simplicity in interactions, and identifying the influential factors. To assess this construct, we planned a two-stage approach: first, we introduced the system to the patient at the beginning of the interview session and showed how it worked in practice; then, we asked the patient to perform the task of entering his or her last visit data to experience how it feels to interact with the system. While working on the different parts of the system, the patient thinks aloud about his or her comments on the program's facilities and interface design.

At the end of the session, we asked participants some open-ended questions in order to assess the user's overall perception of the ease of using the program. The questions were derived from the standardized SUS (System Usability Scale) [19] and PUEU (Perceived Usefulness and Ease of Use) [18] and are available in Multimedia Appendix 2. The results of the data analysis are included in the following section.

Results

In this section, the findings of the data analysis are reported for each phase of the study.

Phase I: Need Analysis

Overall, 46 items were noted by the researcher. Once redundancies were removed, they were classified into nine categories: symptoms, oral drugs, insulin, lab tests, blood glucose test, diet, physical activity, psychological problems, and exposure to new circumstances (Multimedia Appendix 3). According to the purpose of this study, which was to provide explanations through patient records, only four of these thematic categories were found to be associated with the patient's record: symptoms, oral drugs, insulin, and lab tests (Table 1).

Findings from this phase were then used to determine the structure of information in the design phase. The results from this phase revealed the structure of information to be provided for the patient based on their needs.

Table 1. Categorization of patients' questions.

Topic-based category	Concept
Symptoms	Cause of the symptom
	Diagnostic symptoms
	Recommendations
Drugs	Side effects
	Dosage
	Administration and use
Lab tests	Definition
	Normal range
	Patient's status

Phase II: System Design and Implementation

In this research, content quality was controlled not only by health providers but also by patients.

Evaluating Content Quality by Clinicians

Before the system is available to patients, the quality of the content should be ensured. To do this, we asked three experts in the diabetes domain with over 10 years of experience (one endocrinology subspecialist, one general practitioner (GP) who had passed a diabetes course, and one nurse who was a diabetes educator) to evaluate content quality through a checklist (Multimedia Appendix 4). Four criteria were checked: accuracy, simplicity, usefulness, and adequacy. Each aspect is rated between 1 and 5. Table 2 provides mean scores and standard deviations for each aspect.

As it can be observed in Table 2, drugs and diseases received the lowest total scores. Experts believed that the content related to the drug's mechanism of action is unnecessary and difficult for patients. A comparison of mean rating scores indicates that the nurse in charge of diabetes education rated the content quality lower than the physicians. One reason for this can be due to the nurse's better knowledge of patients' learning capabilities, which is a result of more time and closer contact the nurse spent with patients (Multimedia Appendix 4).

Evaluating Content Quality by Patients

Once the health providers approved the content, we inquired about patients' opinions on the quality of information content. This time just the information items related to each patient's record were available to five diabetic patients (three women and two men). The patients had two choices to access the information: on the computer screen or on paper. From the five patients, four requested the paper-based version. Therefore, the information related to every patient's medical record was prepared in a booklet and provided for the patient in the visiting session along with the checklist. The patients had 1 month to read through the booklet and fill out the checklist. Each piece of information was rated on a Likert scale based on four criteria: comprehensibility, practicality, essentiality, and novelty. Moreover, patients were able to add their comments in the explanation section whenever needed.

The patients commented differently on content quality. One reason for such diversity can be due to the varying size and type of information provided for each patient, which itself is a result of the patients' different conditions. The other reasons for this diversity maybe the patients' differing literacy and knowledge levels. Overall, all the participants rated the content quality satisfactory. The patients' rating of comprehensibility, essentiality, and novelty was good or very good.

Phase III: Final System Evaluation

The participants consisted of five diabetes specialists (two trained GPs, two endocrinology subspecialists, and one diabetes nurse) and eight patients who were selected through the convenient sampling method between June and August 2016. Subjects were interviewed for 40 minutes on average. The interviews continued until no new information was added. Table 3 shows the patients' demographic information.

Table 2. Mean scores and standard deviations of health providers' multiple aspects of content quality (n=3).

Item groups	Evaluation aspects				
	Accuracy	Simplicity	Usefulness	Adequacy	Total
Symptoms	4.9 (0.1)	4.7 (0.2)	4.5 (0)	4.6 (0.2)	4.6 (0.1)
Diseases	4.9 (0.2)	4.6 (0.3)	4.4 (0.1)	4 (0.1)	4.4 (0.3)
Drugs	4.8 (0.1)	4.3 (0.3)	4 (0.1)	4.6 (0.2)	4.4 (0.2)
Tests	5 (0)	4.8 (0.1)	5 (0)	5 (0)	4.9 (0.05)
Total score	4.9 (0.1)	4.6 (0.2)	4.4 (0.05)	4.5 (0.1)	

Table 3. Patient participants' demographic information.

Category	Variable	Value
Demographic	Average age (range), in years	63 (57-68)
	Time since diagnosis (mean), in years	12
	Affliction with consequences, in years	4
Medical diet	Drug	3
	Drug + insulin	5
Education	diploma	5
	≥diploma	3
Information sources	Press and TV	5
	Kith and kin	6
	Reading books or educational brochure	4
	Class attendance	4

After the interviews, the audio record was transcribed and then analyzed based on TAM. The analysis involved an identification and categorization of key statements. To do this, each response was analyzed line by line in the related context, and finally the following three themes were found for perceived usefulness: (1) raising patient knowledge and self-awareness; (2) improving patient self-care; and (3) improving doctor-patient interaction.

The specialists' and patients' attitudes toward each of the three themes are presented below.

Raising Patient Knowledge and Self-Awareness

The most remarkable aspect of the system, according to the specialists, is informing patients and making them aware of their medical condition. In this regard, specialist number 1 states:

> *one medical goal for diabetic patients is to make them aware of their disease and equip them with self-care knowledge. Experience has shown that active patients who gain information from different sources manage to control their blood sugar better…*

The patients also maintained that access to their records and its explanations is like carrying a full-time tutor who can be accessed anytime and anywhere. Patient number 2 mentioned that the key to entering the world of diabetes is familiarity with the language of this disease. He continued:

> *…To comprehend texts on diabetes, one needs to have a good command of two languages, the language in which the text is written and the language of diabetes….*

According to the specialists, providing feedback about patient status based on his or her lab test results along with general explanations not only raises patient awareness of one's own conditions but also encourages one to read more on the topic. With this concern, patient number 7 stated:

> *within the 20 years I have been suffering from diabetes, I've read many books and materials. Though I already knew much of what was presented in this system, still when I think this content is especially prepared for me, it becomes more interesting and I tend to read more*

Improving Patient Self-Care

Specialists believed that patient access to his or her record is accompanied by more feelings of responsibility for self-care, which can lead to delayed emergence of complications. In this respect, specialist number 3 maintains:

> *…In my opinion, information tailoring is a sort of attention paid to patients. The short time of the visit does not let every patient's questions be answered and s/he might feel ignored. This method of addressing an individual creates a feeling of importance in him/her, and further motivates reading and practical application. All this makes one believe in one's role in actively managing his/her disease…*

According to his experience of hypoglycemia, patient number 5 stated:

> *I did not know then why I was feeling that way, and did not know what I was supposed to do. The explanations provided in the system both involved preventive advice and treatment recommendations that greatly contributed to lowering the frequency and costs of visiting doctors.*

Improving Doctor-Patient Interaction

Specialists believe that the limited time of a visit will be more effectively used as patient awareness and knowledge improves. One specialist referred back to his own experience of giving instructional leaflets to patients, and believed that patients welcome receiving leaflets from doctors and see it as a sign of respect and attention. He anticipated that patients would read the leaflets, and continued:

> *…the effect of leaflets on changing a patient's behavior and performance can be easily seen in the type of questions asked the following session…*

Patients also believed that accessing their record and receiving explanations helps them better understand the doctor's words. A majority of patients have mentioned that they heard at least

one of the terms in their records from the doctor. However, they did not know what it meant. With this concern, one of them stated:

> *having read the record, I came to know that I mispronounced the names of some drugs…*

Data concerning perceived ease of use was analyzed in a similar fashion. Patients' comments were more in the form of suggestions to improve program facilities and did not mention any problems related to interacting with the system. Textbox 1 shows some of the suggestions made by the specialists about the facilities of the system:

The last item in the Textbox 1 implies that some items might be repeated in the patient's record at different visits. One suggestion was that the information provided should be updated each time. For example, details on blood fat begin with simple basic issues, and once the patient's knowledge is raised over time, more issues that are complex are offered in the following sessions.

Here are some of the issues extracted from patients' comments with respect to perceived ease of use.

Access to all library content: Currently, patients can only view that portion of information that is contained within their record. One suggestion is to provide access to all library content in a tree-like menu. This helps the patient to access his or her record content or search for any other relevant issue that might occur later.

Allowing the patient to enter his or her lab test values and receiving their interpretation: One of the most interesting parts of the program for patients is the analysis of their lab test results along with appropriate advices. At present, the program is designed in a way that the patient-user cannot enter data directly into the system, but can just ask for any data already entered by the doctor. The patients made a request for the possibility of entering lab test results and receiving the interpretation before visiting the doctor. Although this was not considered among the system's primary goals, due to the importance of this feature in better informing the patients and preparing them for the visiting session, it was added to the system as a secondary feature. The mobile-based version of the program, and the possibility of doctor-patient interaction and Q&As were among the other suggestions.

Textbox 1. Suggestions made by specialists about the facilities of the system.

- Considering a possibility of adding to and updating library content by the doctor-user.
- Visualizing test results for a better understanding of less literate patients.
- Taking advantage of audio-visual facilities for more effectiveness.
- Updating the content provided for the patient.

Discussion

Principal Findings

This research dealt with the design, implementation, and evaluation of the first version of a self-describing, tailored, diabetes medical record system. The system design is based on a general approach with no local feature consideration, so it can be adapted to any health care setting with minimum modification, if required. Providing didactic information in the context of medical records is a novel field in information tailoring systems.

According to the body of research, this study is the first step toward providing tailored explanations for diabetic patients in the Persian language. In this research, tailoring is done based on the content of medical records comprising demographic information as well as patients' medical history. Although adding more aspects could improve the quality of information tailoring, there must be a trade-off between the costs of processing more data and raising the quality of the tailored information. Taking records as the basis of identifying relevant topics and interpreting patients' status avoids the data entry burden on the user.

A great body of research into tailored information provision systems is focused on evaluating system effectiveness on patient behavioral and clinical outcomes. A study [20] was conducted on tailored educational materials based on health literacy level and diabetic patients' learning style. The study examined the effectiveness of tailored information on promoting the knowledge of 160 diabetic patients using RCT design. The results obtained revealed that the knowledge of patients who had received tailored educational products was significantly higher than the control group. Another investigation was the "Move more for life" system [21] that produced a tailored educational pamphlet for survivors of breast cancer and evaluated its effect on 330 participants. The results showed that subjects in the intervention group did resistance physical exercises three times more than the control group. A weakness of these investigations is their inadequate elaboration on the features and functions of the tailoring system. In our study, precise and detailed system description helped to fill the gap and facilitate the design of similar systems in the future.

Outcome-based evaluation methods require patients' extensive cooperation and are at risk of unpredicted human and contextual issues, which necessitate considerable time, money, and multiple resources. The researchers' extensive understanding of underlying influential features and the audience's needs and preferences can lead to a more effective system design. Qualitative research provides the researcher with more access to the inner world of an individual, his or her values, preferences, attitudes, and beliefs. In our previous research we suggested that objective criteria do not necessarily reflect actual user satisfaction and that measuring user perceived satisfaction is a more reliable criterion for judging the system's effectiveness in practice [22]. Our study setting was undergoing a transition phase from paper to EMR and, thus, the evaluation approach

in our study belongs to the qualitative domain and includes field observations and qualitative interviews. One of the strengths of this research is the consideration of the target audience's attitudes in all three phases of the study: need analysis, design and implementation, and evaluation.

A majority of computer programs are designed based on the system designers' assumptions about the users' information requirement. This stands as a key reason for the users' low acceptance [12]. Therefore, the first phase of this research dealt with an examination of the target population information needs following field observations.

The final system evaluation was done using a qualitative approach and semistructured interview. In this research, the attitudes of two groups (doctors and patients) were investigated simultaneously, which can be considered as a strength of this study. The results obtained from field observations and interviews have shown that the audiences were ready to accept the system and they had a positive attitude toward using it in practice.

As generalizability is not the purpose of qualitative evaluation methods, we do not make any claim about the applicability of our findings to a wider population or to different contexts. However, comparing the data in Table 3 with the results obtained in study [23], we can consider the study participants both demographically and situationally representative of diabetic patients in Iran. Therefore, we do not expect a dramatic difference in results when recruiting different participants in different contexts.

A substantial *body of literature* has mentioned the important role of users' mental acceptance in increasing the probability of success in transition from paper to an EMR [24]. Hence, findings from our study about the users' readiness and attitudes provide a valuable perspective for health care policy makers in Iran to make more informed decisions in the transfer process.

The unique feature of the system from both the perspectives of patients and specialists is the tailored sense of the information and its effect on motivating one to learn and act better. In this regard, article [22] indicates that the closer the content is to the patient's needs, wants, and perception level, the better that information is digested by the reader and the more practical it becomes.

A comparison of the usability evaluation of the two groups of patients and doctors indicated that patients tended more toward the superficial aspects of the program, whereas doctors' comments were more focused on the procedures and facilities of the program. Together, these two can provide a complete view of the system.

Usability testing conducted on the patients revealed that they preferred to receive the information in the print-based format, which is consistent with the findings of the review article [25]. In this research, those who had received the information printed read the content more than those who received the information in digital form. Similarly, according to the review article [26], despite the provision of audio-visual products and Internet-based programs, paper pamphlets continued to be the most prevalent way of communicating information to the patients. The majority of patients preferred to receive content in print format to study as much as needed in a more relaxed context.

Limitations

The limited number of participants and the limited exposure time to the system are the limitations of this research, which can affect the generalizability of the findings.

The study limitations included incompleteness of some of the patients' data in paper-based records and doctors' illegible handwriting. As this research is a pilot study, technical limitations are not considered at this phase. However, the coverage, speed, and accessibility of the Internet network in Iran are a challenge that needs special attention when implementing the system in reality.

Future Work

Due to the importance of psychological issues in the care of diabetic patients, we suggest considering this aspect in future research in order to achieve more fine-grained tailoring of information. Furthermore, a part of users' suggestions mentioned in the result section can be the basis of future works. Our latter suggestion for future studies is to investigate the long-term effect of the system on patients' outcomes, such as their increased knowledge and self-awareness compared with the usual information provision method.

Acknowledgments

This study is the first author's PhD dissertation, which has been supported by a grant (grant # 950392) from Mashhad University of Medical Sciences Research Council. The authors would like to make special thanks to diabetes specialists from the Endocrine Research Center for their participation in the Delphi Technique.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Two overviews for patient's data representation.

[PDF File (Adobe PDF File), 351KB - medinform_v5i2e10_app1.pdf]

Multimedia Appendix 2

Evaluation Framework and Interview Guide.

[PDF File (Adobe PDF File), 51KB - medinform_v5i2e10_app2.pdf]

Multimedia Appendix 3

Complete list of questions patients asked in observational study.

[PDF File (Adobe PDF File), 213KB - medinform_v5i2e10_app3.pdf]

Multimedia Appendix 4

Content Quality Assessment.

[PDF File (Adobe PDF File), 279KB - medinform_v5i2e10_app4.pdf]

References

1. Lotfi M, Saadati H, Afzali M. Prevalence of diabetes in people aged ≥ 30 years: the results of screen-ing program of Yazd Province, Iran, in 2012. J Res Health Sci 2014;14(1):87-91. [Medline: 24402857]
2. Alharbi N, Alsubki N, Jones S, Khunti K, Munro N, de LS. Impact of information technology–based interventions for type 2 diabetes mellitus on glycemic control: a systematic review and meta-analysis. J Med Internet Res 2016;18(11):e310 [FREE Full text] [doi: 10.2196/jmir.5778] [Medline: 27888169]
3. Schulz P, Nakamoto K. Health literacy and patient empowerment in health communication: the importance of separating conjoined twins. Patient Educ Couns 2013 Jan;90(1):4-11. [doi: 10.1016/j.pec.2012.09.006] [Medline: 23063359]
4. Office for Civil Rights, HHS. Standards for privacy of individually identifiable health information. Final rule; correction of effective and compliance dates. Fed Regist 2001 Feb 26;66(38):12434 [FREE Full text] [Medline: 11706852]
5. Ross S, Lin C. The effects of promoting patient access to medical records: a review. J Am Med Inform Assoc 2003 Mar 01;10(2):129-138. [doi: 10.1197/jamia.M1147]
6. Davis GT, Menon S, Parrish D, Sittig D, Singh H. Patient access to medical records and healthcare outcomes: a systematic review. J Am Med Inform Assoc 2014;21(4):737-741 [FREE Full text] [doi: 10.1136/amiajnl-2013-002239] [Medline: 24154835]
7. Kreuter MW. Tailoring health messages: customizing communication with computer technology. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.
8. Froschl C. IICM. 2005. User modeling and user profiling in adaptive e-learning systems URL: http://http/www.iicm.tugraz.at/thesis/cfroeschl.pdf [accessed 2017-04-11] [WebCite Cache ID 6pewrKzBX]
9. Harrington NG, Noar SM. Reporting standards for studies of tailored interventions. Health Educ Res 2012 Apr;27(2):331-342 [FREE Full text] [doi: 10.1093/her/cyr108] [Medline: 22156230]
10. McRoy SW, Liu-Perez A, Ali SS. Interactive computerized health care education. J Am Med Inform Assoc 1998 Jul 01;5(4):347-356. [doi: 10.1136/jamia.1998.0050347]
11. DiMarco C, Hirst G, Wanner L, Wilkinson J. HealthDoc: customizing patient information and health education by medical condition and personal characteristics. 1995 Aug 12 Presented at: Workshop on Artificial Intelligence in Patient Education; August 12, 1995; Ontario, Canada.
12. Buchanan BG, Moore JD, Forsythe DE, Carenini G, Ohlsson S, Banks G. An intelligent interactive system for delivering individualized information to patients. Artif Intell Med 1995 Apr;7(2):117-154. [Medline: 7647838]
13. Hartson R. The UX Book: Process and Guidelines for Ensuring a Quality User Experience. Waltham, MA: Morgan Kaufmann; 2012.
14. Van Velsen L, Van Der Geest T, Klaassen R, Steehouder M. User-centered evaluation of adaptive and adaptable systems: a literature review. Knowl Eng Rev 2008 Sep 3;23(03):261-281. [doi: 10.1017/S0269888908001379]
15. Delone W. The DeLone and McLean model of information systems success: a ten-year update. JMIS 2003;19(4):9-30. [doi: 10.1080/07421222.2003.11045748]
16. Lee Y, Kozar K, Larsen KR. The technology acceptance model: past, present, and future. CAIS 2003;12(1):50. [doi: 10.1016/j.jbi.2009.07.002]
17. Lee K, Hoti K, Hughes J, Emmerton L. Consumer use of “Dr Google”: a survey on health information-seeking behaviors and navigational needs. J Med Internet Res 2015;17(12):e288. [doi: 10.2196/jmir.4345]
18. Davis FD. Pusefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
19. Brooke J. Meiert. 1996. SUS-A quick and dirty usability scale URL: https://hell.meiert.org/core/pdf/sus.pdf [accessed 2017-04-25] [WebCite Cache ID 6pz6BpW96]
20. Koonce T, Giuse N, Kusnoor S, Hurley S, Ye F. A personalized approach to deliver health care information to diabetic patients in community care clinics*†. J Med Libr Assoc 2015;103(3):123-130. [doi: 10.3163/1536-5050.103.3.004]

21. Short C. Theory-and evidence-based development and process evaluation of the Move More for Life program: a tailored-print intervention designed to promote physical activity among post-treatment breast cancer survivors. Int J Behav Nutr Phys Act 2013;10:124. [doi: 10.1186/1479-5868-10-124]
22. Tara M. IRMA. 2015 Dec 10. Aspects of information tailoring in the 21st century URL: http://www.irma-international.org/viewtitle/112847/ [accessed 2017-04-25] [WebCite Cache ID 6pz6WjSfd]
23. Veghari G, Sedaghat M, Joshaghani H, Hoseini SA, Niknezad F, Angizeh A, et al. Association between socio-demographic factors and diabetes mellitus in the north of Iran: a population-based study. Int J Diabetes Mellit 2010 Dec;2(3):154-157. [doi: 10.1016/j.ijdm.2010.09.001]
24. Abdekhoda M, Ahmadi M, Dehnad A, Hosseini A. Information technology acceptance in health information management. Methods Inf Med 2014;53(1):14-20. [doi: 10.3414/ME13-01-0079] [Medline: 24248003]
25. Pal K, Eastwood S, Michie S, Farmer A, Barnard M, Peacock R, et al. Computer-based interventions to improve self-management in adults with type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2014 Jun;37(6):1759-1766. [doi: 10.2337/dc13-1386] [Medline: 24855158]
26. Kenny T, Wilson R, Purves I, Clark JS, Newton L, Newton D, et al. A PIL for every ill? Patient information leaflets (PILs): a review of past, present and future use. Fam Pract 1998 Oct;15(5):471-479. [Medline: 9848435]

Abbreviations

EMR: electronic medical records
GP: general practitioner
TAM: technology acceptance model

Edited by CL Parra-Calderón; submitted 21.10.16; peer-reviewed by S Baptista, S Tao, A Martinez-Millana; comments to author 03.01.17; revised version received 11.02.17; accepted 05.03.17; published 02.05.17.

Please cite as:
Kamel Ghalibaf A, Mazloum Khorasani Z, Gholian-Aval M, Tabesh H, Tara M
Design, Implementation, and Evaluation of Self-Describing Diabetes Medical Records: A Pilot Study
JMIR Med Inform 2017;5(2):e10
URL: http://medinform.jmir.org/2017/2/e10/
doi:10.2196/medinform.6862
PMID:28465280

Original Paper

Computerized Childbirth Monitoring Tools for Health Care Providers Managing Labor: A Scoping Review

Michael S Balikuddembe[1,2,3], MMED, Dip SRHR; Nazarius M Tumwesigye[4], MSc (Medical statistics), PhD; Peter K Wakholi[5], MSc, PhD; Thorkild Tylleskär[1], MD, PhD

[1]Center for International Health, University of Bergen, Bergen, Norway

[2]Department of Epidemiology and Biostatistics, Makerere University, Kampala, Uganda

[3]Department of Obstetrics & Gynaecology, Mulago National Referral and Teaching Hospital, Kampala, Uganda

[4]School of Public Health, Department of Epidemiology & Biostatistics, Makerere University, Kampala, Uganda

[5]School of Computing & Informatics Technology, Makerere University, Kampala, Uganda

Corresponding Author:
Michael S Balikuddembe, MMED, Dip SRHR
Department of Epidemiology and Biostatistics
Makerere University
SPH Bldg. Mulago
Kampala,
Uganda
Phone: 256 99876602
Email: balikuddembem@gmail.com

Abstract

Background: Proper monitoring of labor and childbirth prevents many pregnancy-related complications. However, monitoring is still poor in many places partly due to the usability concerns of support tools such as the partograph. In 2011, the World Health Organization (WHO) called for the development and evaluation of context-adaptable electronic health solutions to health challenges. Computerized tools have penetrated many areas of health care, but their influence in supporting health staff with childbirth seems limited.

Objective: The objective of this scoping review was to determine the scope and trends of research on computerized labor monitoring tools that could be used by health care providers in childbirth management.

Methods: We used key terms to search the Web for eligible peer-reviewed and gray literature. Eligibility criteria were a computerized labor monitoring tool for maternity service providers and dated 2006 to mid-2016. Retrieved papers were screened to eliminate ineligible papers, and consensus was reached on the papers included in the final analysis.

Results: We started with about 380,000 papers, of which 14 papers qualified for the final analysis. Most tools were at the design and implementation stages of development. Three papers addressed post-implementation evaluations of two tools. No documentation on clinical outcome studies was retrieved. The parameters targeted with the tools varied, but they included fetal heart (10 of 11 tools), labor progress (8 of 11), and maternal status (7 of 11). Most tools were designed for use in personal computers in low-resource settings and could be customized for different user needs.

Conclusions: Research on computerized labor monitoring tools is inadequate. Compared with other labor parameters, there was preponderance to fetal heart monitoring and hardly any summative evaluation of the available tools. More research, including clinical outcomes evaluation of computerized childbirth monitoring tools, is needed.

(JMIR Med Inform 2017;5(2):e14) doi:10.2196/medinform.6959

KEYWORDS

childbirth; obstetric labor; fetal monitoring; medical informatics applications; systematic review

Introduction

In 2015, an estimated 303,000 women died from pregnancy-related complications such as excessive bleeding, obstructed labor, and infections [1,2]. The obstructed labor complex directly contributes to 6-10% of the maternal deaths, in addition to contributing to other diseases for the mother and the baby [3,4].

Proper monitoring of the labor and delivery process with appropriate action based on findings is one of the keys to the prevention of pregnancy-related diseases and deaths [5,6]. Labor monitoring includes three main areas, namely fetal conditions, labor progress, and maternal conditions. The fetal parameters include fetal heart rate and amniotic fluid color, whereas labor progress is tracked through cervical dilation, uterine contractions, and fetal descent. The parturient's condition is monitored by her blood pressure, temperature, urine, and mental state. The monitoring in many low-resource settings is hampered by the lack of user-friendly tools for labor management, limited access to evidence-based clinical guidelines for the providers of maternal health services, maternity provider factors, weak referral networks, and limited health financing [7].

Since 1994, the paper partograph has been promoted by the World Health Organization (WHO) as the standard labor monitoring tool [8], but to date its use is still poor in many low-resource settings due to many user and usability challenges [1,9-11]. To address these challenges, scientists in maternal health called for improvements of the partograph [1,12,13]. The WHO called for the development and evaluation of pragmatic electronic health (eHealth) solutions to health challenges [14].

Noteworthy, mobile health (mHealth) was embraced in many settings, especially chronic conditions with concomitant improvement in medical care [15,16]. It was hoped that next-generation system innovations could improve the quality of care during childbirth and reduce maternal deaths [17,18]. However, there seems to be a paucity of papers on computerized labor monitoring tools as is with mHealth in general.

We had a notion that the responses to various calls for better labor monitoring tools are still poor. Therefore, we set out to determine the volume and scope of research on computerized monitoring tools that can be used by health care providers in childbirth management.

Methods

We undertook a scoping review, as defined elsewhere [19,20], to assess the reactions to the WHO call for labor monitoring tools (computerized or otherwise) with the potential of being more acceptable to the stakeholders in maternity services. Tricco et al (2016) reiterate the purpose of scoping reviews as, "…to present a broad overview of the evidence pertaining to a topic, irrespective of study quality, and are useful when examining areas that are emerging, to clarify key concepts and identify gaps" [19]. In June and July 2016, we searched PubMed and Google Scholar databases for peer-reviewed and gray literature. The search was supplemented by manual searches in Google search engine and ResearchGate online repositories for other papers meeting the selection criteria.

The inclusion criteria were a paper written in English language, addressing an aspect of a computerized or mobile labor monitoring tool, for use by maternity service providers, and written between January 1, 2006 to May 31, 2016. We excluded literature on tools for use primarily by expectant parents and the numerous apps for nonprofessional use such as contraction monitoring at home or fetal growth monitoring.

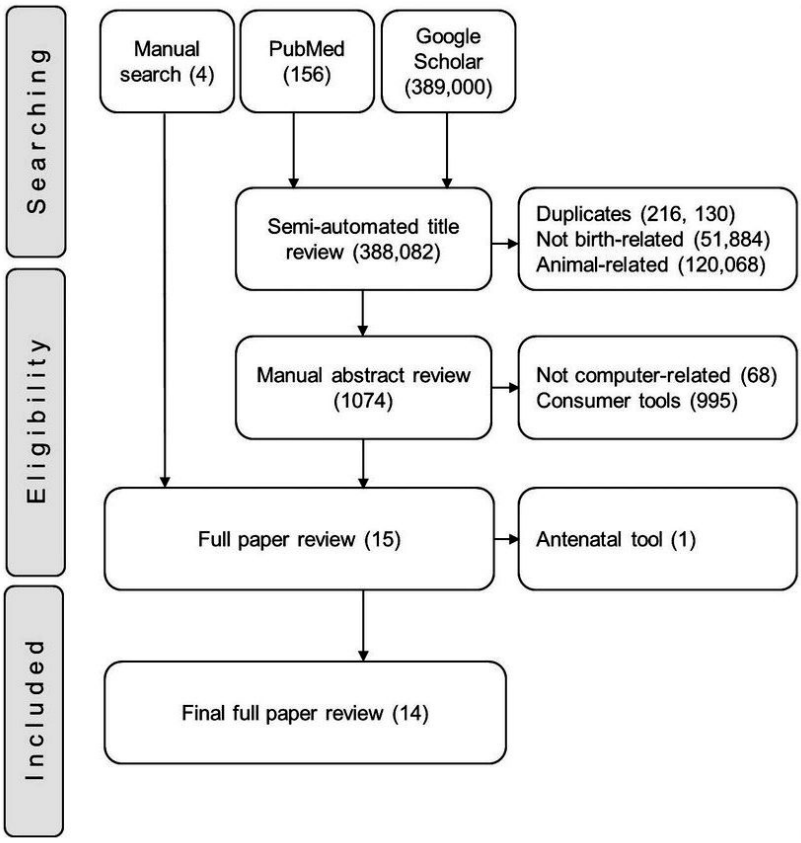
Our key terms in the search included "labor," "monitoring," "computer," "mobile," "tool," "provider," "delivery," and "birth." We combined them in various ways to get search strings. An example of such a string is "With all 'labor monitoring' + plus at least one of 'computer$ tool$ mobile provider$ delivery birth - consumer' (anywhere in article)." At the title review stage, we combined the term "labor" with one or more of the other terms to get relevant papers.

Data collection started with the individual researcher or research assistant identifying and screening papers for allocated years. We then entered the search terms and filtered the results according to the desired years of publication. We sorted the results in ascending years to ease tracking of the viewed Web pages. Each collector exported the identified papers into Mendeley-1.16.1 reference manager (by Elsevier) and used it to remove duplicates.

The data collector imported the Mendeley (by Elsevier) output into Google Scholar or PubMed and applied more specific filters to the titles. The filters helped eliminate the nonhuman birth-related and non-computer tools. The subsequent titles were saved to an online library for further screening. Abstracts to the saved titles were downloaded and qualitatively analyzed to determine the target users. Uncertainty about the eligibility of an article was consensually resolved based on the selection criteria.

Full papers to abstracts with all eligibility criteria were retrieved. Manually identified abstracts or papers that were not part of the controlled search were added to the pool. All papers were scrutinized against all selection criteria. The remaining papers were included in the quantitative final analysis. These steps are summarized, as shown in Figure 1, to mimic the preferred reporting items for systematic reviews and meta-analyses (PRISMA). Each paper was read to decipher the focus, developmental stage, and computing platform of the labor monitoring tool.

Figure 1. The steps of paper selection depicted in a PRISMA flow diagram. The number of articles included or excluded at each step is shown in brackets.



Results

Volume of Relevant Research

In the preliminary literature search, we retrieved 389,000 titles from Google Scholar and 156 from PubMed (Figure 1). These were imported into Mendeley 1.16.1 (by Elsevier) and duplicates removed. The titles were semiautomatically reviewed to remove papers about nonhuman or nonbirth related tools, which resulted in the exclusion of 388,082 entries. During the screening of abstracts for the remaining 1074 papers, we eliminated papers about tools that did not incorporate computers during labor monitoring use to leave 11 abstracts. Four abstracts identified in the manual literature search were added to the remaining 11 to get 15 abstracts used to identify papers for full review. At the full paper review stage, one paper (the Bacis program study by Horner in South Africa) was excluded because its subject tool was not designed for use during labor, which left 14 papers with all inclusion criteria for final analysis, as shown in Table 1.

Stage of Tool Development

As shown in Table 2, 4 out of 14 analyzed papers addressed labor monitoring tools still at the planning phase. We analyzed 7 papers about tools that had reached the implementation stage. The authors of the remaining 2 papers addressed the cost and impact on clinic workflow evaluation of one tool, QUALMAT eCDSS [21].

Five tools that were in the clinical trial or field testing [22-25] phase were classified under the implementation stage of development. Three tools had undergone formative assessment in form of user or cost evaluations [23,26,27]. We retrieved literature on a tool that was designed and tested in 2007, but we did not get publications on its advancement. One tool had a summative evaluation of its cost and impact on clinic workflow [28]. We did not get any tool with definitive summative evaluation, that is, pragmatic clinical outcomes (morbidity or mortality) studies.

Focus of Labor Monitoring Tool

The main parameters of focus in the analyzed papers were monitoring all labor parameters (7 papers) and fetal heart (3 papers). Four tools were designed to monitor one of three labor monitoring sections, whereas the rest could monitor multiple sections, as shown in Table 3. Authors addressing multiple labor monitoring sections were chiefly concerned with electronic versions of the partograph computing abilities. Therefore, their innovations potentially addressed the 15 parameters on a partograph. Those on fetal status focused on fetal heart monitoring with a cardiotocogram (CTG). The authors who addressed labor progress only specifically targeted cervical dilation and descent or station of a cephalic fetus.

Table 1. List of papers included in the final analysis and the name of study tools.

Paper (Reference number)	Year[a]	Name of tool
"Quality of prenatal and maternal care: Bridging the know-do gap" (QUALMAT study): an electronic clinical decision support system for rural sub-Saharan Africa [21]	2013	QUALMAT eCDSS[b]
The Moyo fetal heart rate monitor [22]	2014	Moyo monitor
mLabour: design and evaluation of a mobile partograph and labor ward management application [23]	2016	mLabour
Continuous monitoring of cervical dilatation and fetal head station during labor [24]	2007	Computerized labor-monitor
The design and implementation of the PartoPen maternal health monitoring system [25]	2013	PartoPen
Improving maternal labor monitoring in Kenya using digital pen technology: a user evaluation [26]	2012	PartoPen
Cost-effectiveness of a clinical decision support system in improving maternal health care in Ghana [27]	2015	QUALMAT eCDSS
Impact of an electronic clinical decision support system on workflow in antenatal care: the QUALMAT eCDSS in rural health care facilities in Ghana and Tanzania [28]	2015	QUALMAT eCDSS
A mobile multi-agent information system for ubiquitous fetal monitoring [29]	2014	Fetal IMAIS
A study of an intelligent system to support decision making in the management of labour using the cardiotocograph–the infant study protocol [30]	2016	INFANT
The development of a simplified, effective, labour monitoring-to-action (SELMA) tool for better outcomes in labour difficulty (BOLD): Study protocol [31]	2015	SELMA
Life curve mobile application: an easier alternative to paper partograph [32]	2015	Life curve
ePartogram: a mobile decision support tool to address labor complications [33]	2013	ePartogram
Another set of eyes: Remote fetal monitoring surveillance aids the busy labor and delivery unit [34]	2010	ANGEL shield

[a]Year of publication.

[b]eCDSS: electronic clinical decision support system.

Table 2. Stages of development for the computerized labor monitoring tools.

Tool	Stage of development[a]				
	Plan	Designing	Implementing	Formative evaluation	Summative evaluation
QUALMAT eCDSS[b]	y[c]	y	y	y	z
Moyo monitor	y	y	z[c]	z	x
mLabour	y	y	z	y	x
Computerized labor-monitor	y	y	z	x	x
PartoPen	y	y	y	y	x
Fetal IMAIS	y	x[c]	x	x	x
INFANT	y	x	x	x	x
SELMA	y	x	x	x	x
Life curve	y	x	x	x	x
ePartogram	y	y	z	x	x
ANGEL shield	y	y	y	z	x

[a]Based on the five stages of software development.

[b]eCDSS: electronic clinical decision support system.

[c]x: stage not reached, y: completed stage, z: stage incomplete.

Table 3. Parameters on modified WHO partograph noted in the capability of computerized tools.

Parameter on WHO partograph	Computerized tool to monitor parameter										
	21[a]	22[b]	23[c]	24[d]	25[e]	29[f]	30[g]	31[h]	32[i]	33[j]	34[k]
Fetal heart	y[l]	y	y		y	y	y	y	y	y	y
Cervix opening	y		y	y	y		y	y	y	y	
Descent of leading part	y		y	y	y		y	y	y	y	
Amniotic fluid	y		y		y		y	y	y	y	
Molding of head	y		y		y		y	y	y	y	
Uterine contractions	y		y	y	y		y	y	y	y	
Maternal pulse	y		y		y		y	y	y	y	
Maternal blood pressure	y		y		y		y	y	y	y	
Maternal temperature	y		y		y		y	y	y	y	
Urine protein	y		y		y		y	y	y	y	
Urine acetone	y		y		y		y	y	y	y	
Urine volume	y		y		y		y	y	y	y	
Time of membrane rupture	y		y		y		y	y	y	y	
Drug given	y		y		y		y	y	y	y	
Clinical diagnosis suggestion	y	y	y		y		y	y	y	y	

[a]QUALMAT eCDSS.
[b]Moyo monitor.
[c]mLabour.
[d]Computerized labor-monitor.
[e]PartoPen.
[f]Fetal IMAIS.
[g]INFANT.
[h]SELMA.
[i]Life curve.
[j]ePartogram.
[k]ANGEL shield.
[l]Parameter can be monitored with the tool.

Table 4. Computing platforms and adaptability for the labor monitoring tools.

Tools	Computing Platform				Adaptable
	Portability	Network environment	Operating systems	Software base	
QUALMAT	Laptop and desktop	Stand-alone	MS Windows	Java	Yes
Moyo monitor	Mobile	Stand-alone	Customized	Undisclosed	No
mLabour	Mobile	Client-server	Android	Application	Yes
Computerized labor-monitor	Mobile and desktop	Stand-alone	Nonspecific	Ultrasound waves	Unknown
PartoPen	Mobile	Stand-alone	Livescribe pen	LiveCode “Penlet”	No
Fetal IMAIS	Mobile & desktop	Client-server, wireless	MS Windows	Java	Yes
INFANT	Mobile and desktop	Offline	MS Windows	Undisclosed	Unknown
SELMA	Mobile and desktop	Offline	Not applicable	Undisclosed	Unknown
Life curve	Mobile	Client-server	Android	Application	Yes
ePartogram	Mobile	Client-server	Android	Undisclosed	Unknown
ANGEL shield	Desktop	Client-server	MS Windows	Undisclosed	Yes

All tools were intended for clinical diagnosis support based on algorithms. The least basic was the capability to take measurements or give alerts and reminders to maternity care providers [24,29,34]. However, 8 of 11 tools could also provide diagnosis and action suggestions to the user. The proposed SELMA tool could use machine learning models to predict diagnosis and outcomes for different contexts of use.

Computing Platforms for the Labor Monitoring Tools

As shown in Table 4, the tools were mostly usable on existing personal computer hardware, especially mobile gadgets such as phones, laptops, and desktop computers. Majority communicated through client-server networks, and 4 of 11 (36%) used stand-alone computers. Microsoft Windows was the most commonly used operating system in desktops, and Android was used in most mobile systems. One tool (PartoPen) uses custom-made software, whereas the application-based tools were developed in a Java environment. For half of the tools, the authors did not specify the software framework used in development. Of 11 tools, only 4 were reported as customizable to suit a user's context of work.

The authors of the analyzed papers planned for adaptable tools. The intended context of tool use was stated as low-resource settings apart from the ANGEL shield. This computer program was made for and operated in a university hospital, but there were plans of rolling it out to nearby rural health centers. For 5 of 11 tools, it was reported that they could be customized to different contexts of use and even more functions added, where necessary.

Discussion

Principal Findings

In this review, out of over 380,000 papers, 14 qualified for the final analysis. They represented studies of 11 computerized tools capable of aiding health workers in maternity care. All labor parameters could be monitored by 7 of the 11 tools upon implementation. Only one tool had summative evaluations, which included cost and indicator studies. We did not find evidence of morbidity or mortality evaluation for any tool. Most tools used open source software that was also adaptable to common computers.

In this study, we chose to conduct a scoping review due to an ostensibly low volume of systematic evaluations and publications but with potentially more works on the subject. To this effect, many papers on pregnancy care applications were identified, although they were excluded from analysis for failure to meet other inclusion criteria. PubMed and Google Scholar formed the basis for the main search results due to their popularity among authors and a wide coverage of subjects. They were augmented with a manual search that indeed yielded more papers [22,23,32,33] included in the final analysis. The inclusion period of 2006 to 2016 was chosen to encompass the 5 years before and after the 2011 WHO call to evaluate all design and scale-up stages in eHealth [14], including e-labor monitoring tools.

Many papers were identified, which echo the findings of Kortteisto et al (2014), that is, mHealth is widespread [16]. However, similar to WHO and International Confederation of Midwives concerns in 2011 [13,14], only a handful addressed maternal labor and delivery monitoring. Moreover, as Hall et al found [17], the authors reported on tools that were in developmental stages, protocol preparation to field testing, without the definitive summative evaluation (pragmatic clinical outcomes studies) needed in health care research [35].

This state of affairs may be due to the human-intensive nature and high litigation potential of labor monitoring events that designing and testing a reliable computerized labor tool calls for more effort than is needed for an average medical condition. Another factor that could deter innovators is the difficulty in definitive summative evaluation of the tools in light of many confounders of labor outcomes. A similar situation in 2011 could have led the WHO to call for better research and evaluation of mHealth [14] such as labor monitoring tools. One should also be cognizant of the diverse labor monitoring contexts both within and without health facilities and communities, which were highlighted in the 2016 Lancet maternal health series [36,37]. However, the trend of research seems to be in a positive direction. From the works analyzed in this paper, the majority of the papers are authored after the WHO call to action, and all the documented evaluations were published after the call.

After the data collection for this review, another call to action on improving quality of maternity and newborn care was sounded through the Lancet maternal health series [37]. Furthermore, the potential of mHealth to help out is anticipated, with innovators urged not to be stifled by the fear of liability and litigation but to provide evidence-based tools for woman-centered care [18,38]. This series reinforced the findings of this review and stressed the need for further research for context-specific mobile childbirth monitoring tools.

Regarding the focus of the tools, it was obvious that the three main sections—fetal, labor progress, and maternal states—of labor monitoring received unequal attention. Monitoring the fetal condition, especially the fetal heart, was most researched, perhaps due to the discovery of cardiotocogram (CTG) that is efficacious in detecting fetal distress. CTG is sonoelectric and improving it through computerization was easier than inventing tools as was necessary for the labor progress. On the other hand, cervical dilation and maternal conditions are subjectively dependent on provider skills, and sociocultural or religious norms. Hence, the diversity of these conditions could hamper the design, testing, and development of widely acceptable tools to monitor labor progress and maternal conditions. With the recognition of the diversity of contexts of use but limited resources, it is prudent that generic tools are designed and developed or adapted for specific contexts [39].

The operating system platforms were generic in about half of the tools perhaps to cut development costs. This is also good for the diverse hardware that is increasingly portable. On the other hand, the programming software base was mostly undisclosed. This could be due to the early stages of development. Moreover, the authors were not yet committed to a specific program. However, the generic Java development environment was used in most specified cases. This would

suggest that even the rest are more likely to use it in tool development.

Limitations

The sources of data for this review were not exhaustive of all literature (eg, papers not written in English), and as such, we could have omitted some papers on the computerized labor monitoring tools. This is a drawback, but the main sources are wide enough for an adequate sample and similar reviewers [17,20,40] increasingly adopt this approach. We also started the search with broad terms and narrowed them down as we saved different papers for subsequent analyses.

Conclusions

In conclusion, the scope and volume of research on computerized maternal labor monitoring tools is likely to be narrow and small. Fetal heart monitoring seems to dominate over other labor management parameters. Most tools are designed to use affordable computing platforms, but there is hardly any summative evaluation of the available tools. This may imply a slow response to the call for developing and evaluating computerized labor monitoring tools that could reduce labor-related disease. Further research, including clinical outcomes studies and publication of results, is needed on computerized tools for use in comprehensive childbirth monitoring.

Acknowledgments

We thank the Norwegian Agency for Development (NORAD) through the HITRAIN project and University of Bergen, UiB, for sponsoring this work. We thank the Association of Obstetricians & Gynecologists of Uganda secretariat for the support during the literature search and manuscript preparation. This work was supported by NORAD's NORHED program through the HITRAIN project. However, the program had no role in determining the study design, data collection and analysis, or in the interpretation of results and writing of the manuscript.

Authors' Contributions

All authors contributed to the review protocol. MB wrote the review protocol and participated in data collection and analysis. TT participated in data collection and analysis. All authors participated in manuscript preparation and approval of its final copy.

Conflicts of Interest

None declared.

References

1. Graham W, Woodd S, Byass P, Filippi V, Gon G, Virgo S, et al. Diversity and divergence: the dynamic burden of poor maternal health. Lancet 2016 Dec 29;388(10056):2164-2175. [doi: 10.1016/S0140-6736(16)31533-1] [Medline: 27642022]
2. Say L, Chou D, Gemmill A, Tunçalp Ö, Moller A, Daniels J, et al. Global causes of maternal death: a WHO systematic analysis. Lancet Glob Health 2014 Jun;2(6):e323-e333 [FREE Full text] [doi: 10.1016/S2214-109X(14)70227-X] [Medline: 25103301]
3. Ollerhead E, Osrin D. Barriers to and incentives for achieving partograph use in obstetric practice in low- and middle-income countries: a systematic review. BMC Pregnancy Childbirth 2014 Aug 16;14:281 [FREE Full text] [doi: 10.1186/1471-2393-14-281] [Medline: 25132124]
4. Balikuddembe MS, Byamugisha JK, Sekikubo M. Impact of decision to operation interval on pregnancy outcomes among mothers who undergo emergency caesarean section at Mulago hospital. J US China Med Sci 2011 Sep;8(9):571-575 [FREE Full text]
5. Ogwang S, Karyabakabo Z, Rutebemberwa E. Assessment of partogram use during labour in Rujumbura Health Sub District, Rukungiri District, Uganda. Afr Health Sci 2009 Aug 1;9(Suppl 1):S27-S34 [FREE Full text] [Medline: 20589158]
6. Windrim R, Seaward PG, Hodnett E, Akoury H, Kingdom J, Salenieks ME, et al. A randomized controlled trial of a bedside partogram in the active management of primiparous labour. J Obstet Gynaecol Can 2007 Jan;29(1):27-34. [Medline: 17346475]
7. World Bank. Worldbank. 2012. Uganda makes progress on maternal health but serious challenges remain URL: http://www.worldbank.org/en/news/feature/2012/10/23/uganda-makes-progress-on-maternal-health-but-serious-challenges-remain [WebCite Cache ID 6llF5wtL2]
8. Lavender T, Hart A, Smyth RM. Effect of partogram use on outcomes for women in spontaneous labour at term. Cochrane Database Syst Rev 2012 Aug 15(8):CD005461 [FREE Full text] [doi: 10.1002/14651858.CD005461.pub3] [Medline: 22895950]
9. Fistula Care and Maternal Health Task Force. Bettercaretogether. New York: EngenderHealth/Fistula Care; 2012. Revitalizing the partograph: does the evidence support a global call to action? New YorkngenderHealth/Fistula Care 2012; URL: http://www.bettercaretogether.org/sites/default/files/resources/EngenderHealth-Fistula-Care-Partograph-Meeting-Report-9-April-12.pdf [WebCite Cache ID 6llFgp9DG]

10. Okokon IB, Oku AO, Agan TU, Asibong UE, Essien EJ, Monjok E. An evaluation of the knowledge and utilization of the partogragh in primary, secondary, and tertiary care settings in calabar, South-South Nigeria. Int J Family Med 2014;2014:105853 [FREE Full text] [doi: 10.1155/2014/105853] [Medline: 25298892]
11. Weerasekara D. Usefulness of a partograph to improve outcomes: scientific evidence. Sri Lanka J Obstet & Gynae 2014;36(2):29-31. [doi: 10.4038/sljog.v36i2.7449]
12. Orhue A, Ande A, Aziken M. Barriers to the universal application of the partograph for labour management: a review of the issues and proposed solutions. Trop J Obstet Gynaecol 2011 Oct;28(1):32-42.
13. International Confederation of Midwives. Partograph side meeting position statement. 2014 Presented at: ICM Congress; 2014; Prague URL: http://www.mchip.net/sites/default/files/ICM%20Partograph%20Meeting%20Report.pdf
14. Amoroso C, Arango J, Bailey C, Blaya J, Fraser H, Gakuba R, et al. A call to action on global eHealth evaluation. 2011 Presented at: Bellagio eHealth Evaluation Meeting; 2011; Bellagio URL: https://www.ghdonline.org/uploads/The_Bellagio_eHealth_Evaluation_Call_to_Action-Release.docx
15. Blaya JA, Fraser HS, Holt B. E-health technologies show promise in developing countries. Health Aff (Millwood) 2010 Feb;29(2):244-251 [FREE Full text] [doi: 10.1377/hlthaff.2009.0894] [Medline: 20348068]
16. Kortteisto T, Raitanen J, Komulainen J, Kunnamo I, Mäkelä M, Rissanen P, EBMeDS (Evidence-Based Medicine electronic Decision Support) Study Group. Patient-specific computer-based decision support in primary healthcare--a randomized trial. Implement Sci 2014;9:15 [FREE Full text] [doi: 10.1186/1748-5908-9-15] [Medline: 24444113]
17. Hall CS, Fottrell E, Wilkinson S, Byass P. Assessing the impact of mHealth interventions in low- and middle-income countries--what has been shown to work? Glob Health Action 2014;7:25606 [FREE Full text] [Medline: 25361730]
18. Kruk ME, Kujawski S, Moyer CA, Adanu RM, Afsana K, Cohen J, et al. Next generation maternal health: external shocks and health-system innovations. Lancet 2016 Dec 05;388(10057):2296-2306. [doi: 10.1016/S0140-6736(16)31395-2] [Medline: 27642020]
19. Tricco AC, Lillie E, Zarin W, O'Brien K, Colquhoun H, Kastner M, et al. A scoping review on the conduct and reporting of scoping reviews. BMC Med Res Methodol 2016 Feb 09;16:15 [FREE Full text] [doi: 10.1186/s12874-016-0116-4] [Medline: 26857112]
20. Sharafoddini A, Dubin JA, Lee J. Patient similarity in prediction models based on health data: a scoping review. JMIR Med Inform 2017 Mar 03;5(1):e7 [FREE Full text] [doi: 10.2196/medinform.6730] [Medline: 28258046]
21. Blank A, Prytherch H, Kaltschmidt J, Krings A, Sukums F, Mensah N, et al. "Quality of prenatal and maternal care: bridging the know-do gap" (QUALMAT study): an electronic clinical decision support system for rural Sub-Saharan Africa. BMC Med Inform Decis Mak 2013 Apr 10;13:44 [FREE Full text] [doi: 10.1186/1472-6947-13-44] [Medline: 23574764]
22. Laerdal. Laerdalglobalhealth. 2014. Moyo fetal heart rate monitor URL: http://www.laerdalglobalhealth.com/doc/2555/Moyo-Fetal-Heart-Rate-Monitor [accessed 2016-11-04] [WebCite Cache ID 6llGiof3C]
23. Schweers J, Khalid M, Underwood H, Bishnoi S, Chhugani M. mLabour: design and evaluation of a mobile partograph and labor ward management application. Procedia Eng 2016;159:35-43. [doi: 10.1016/j.proeng.2016.08.061]
24. Sharf Y, Farine D, Batzalel M, Megel Y, Shenhav M, Jaffa A, et al. Continuous monitoring of cervical dilatation and fetal head station during labor. Med Eng Phys 2007 Jan;29(1):61-71. [doi: 10.1016/j.medengphy.2006.01.005] [Medline: 16495119]
25. Underwood H, Sterling SR, Bennett JK. The design and implementation of the partopen maternal health monitoring system. In: ACM Digital Library.: ACM; 2013 Presented at: Proceedings of the 3rd ACM Symposium on Computing for Development; January 11 - 12, 2013; Bangalore, India p. 1-10. [doi: 10.1145/2442882.2442893]
26. Underwood H, Sterling SR, Bennett JK. Improving maternal labor monitoring in Kenya using digital pen technology: a user evaluation. : IEEE; 2012 Dec Presented at: Global Humanitarian Technology Conference (GHTC), 2012 IEEE; Oct 21 - 24, 2012; Seattle, WA, USA URL: http://ieeexplore.ieee.org/document/6387062/ [doi: 10.1109/GHTC.2012.77]
27. Dalaba MA, Akweongo P, Aborigo RA, Saronga HP, Williams J, Blank A, et al. Cost-effectiveness of clinical decision support system in improving maternal health care in Ghana. PLoS One 2015;10(5):e0125920 [FREE Full text] [doi: 10.1371/journal.pone.0125920] [Medline: 25974093]
28. Mensah N, Sukums F, Awine T, Meid A, Williams J, Akweongo P, et al. Impact of an electronic clinical decision support system on workflow in antenatal care: the QUALMAT eCDSS in rural health care facilities in Ghana and Tanzania. Glob Health Action 2015;8:25756 [FREE Full text] [Medline: 25630707]
29. Su C, Chu T. A mobile multi-agent information system for ubiquitous fetal monitoring. Int J Environ Res Public Health 2014 Jan 02;11(1):600-625 [FREE Full text] [doi: 10.3390/ijerph110100600] [Medline: 24452256]
30. Brocklehurst P. A study of an intelligent system to support decision making in the management of labour using the cardiotocograph - the INFANT study protocol. BMC Pregnancy Childbirth 2016 Jan 20;16:10 [FREE Full text] [doi: 10.1186/s12884-015-0780-0] [Medline: 26791569]
31. Souza JP, Oladapo OT, Bohren MA, Mugerwa K, Fawole B, Moscovici L, WHO BOLD Research Group. The development of a Simplified, Effective, Labour Monitoring-to-Action (SELMA) tool for Better Outcomes in Labour Difficulty (BOLD): study protocol. Reprod Health 2015 May 26;12(1). [doi: 10.1186/s12978-015-0029-4] [Medline: 26006758]
32. Hasan J, Hossain T, Arafat Y, Karim G, Ahmed Z, Mafi MS. Mobile application can be an effective tool for reduction of maternal mortality. International Journal of Perceptions in Public Health 2017 Mar;1(2):92-94.

33. Dekel S, Mandel B. jhpiego. ePartogram; 2013. A mobile decision support tool to address labour complications URL: https://www.jhpiego.org/video/epartogram-a-mobile-decision-support-tool/ [accessed 2016-11-04] [WebCite Cache ID 6llI8PuMm]
34. Williams D, Ivey T, Benton T, Rhoads S. Another set of eyes: Remote fetal monitoring surveillance aids the busy labor and delivery unit. J Obstet Gynecol Neonatal Nurs 2010 Sep;39:S41. [doi: 10.1111/j.1552-6909.2010.01119_30.x]
35. Day K, Kenealy TW, Sheridan NF. Should we embed randomized controlled trials within action research: arguing from a case study of telemonitoring. BMC Med Res Methodol 2016 Jun 08;16:70 [FREE Full text] [doi: 10.1186/s12874-016-0175-6] [Medline: 27277940]
36. Campbell OMR, Calvert C, Testa A, Strehlow M, Benova L, Keyes E, et al. The scale, scope, coverage, and capability of childbirth care. Lancet 2016 Dec 29;388(10056):2193-2208. [doi: 10.1016/S0140-6736(16)31528-8] [Medline: 27642023]
37. Koblinsky M, Moyer CA, Calvert C, Campbell J, Campbell OMR, Feigl AB, et al. Quality maternity care for every woman, everywhere: a call to action. Lancet 2016 Dec 05;388(10057):2307-2320. [doi: 10.1016/S0140-6736(16)31333-2] [Medline: 27642018]
38. Shaw D, Guise J, Shah N, Gemzell-Danielsson K, Joseph KS, Levy B, et al. Drivers of maternity care in high-income countries: can health systems support woman-centred care? Lancet 2016 Dec 05;388(10057):2282-2295. [doi: 10.1016/S0140-6736(16)31527-6] [Medline: 27642026]
39. English LL, Dunsmuir D, Kumbakumba E, Ansermino JM, Larson CP, Lester R, et al. The paediatric risk assessment (para) mobile app to reduce postdischarge child mortality: design, usability, and feasibility for health care workers in Uganda. JMIR Mhealth Uhealth 2016 Feb 15;4(1):e16 [FREE Full text] [doi: 10.2196/mhealth.5167] [Medline: 26879041]
40. Daudt HM, van Mossel C, Scott SJ. Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. BMC Med Res Methodol 2013;13:48 [FREE Full text] [doi: 10.1186/1471-2288-13-48] [Medline: 23522333]

Abbreviations

CTG: cardiotocogram
eCDSS: electronic clinical decision support system
eHealth: electronic health
mHealth: mobile health
PRISMA: preferred reporting items for systematic reviews and meta-analyses
WHO: World Health Organization

Edited by G Eysenbach; submitted 07.11.16; peer-reviewed by F Sukums, A Blank; comments to author 04.02.17; revised version received 11.02.17; accepted 11.04.17; published 15.06.17.

Please cite as:
Balikuddembe MS, Tumwesigye NM, Wakholi PK, Tylleskär T
Computerized Childbirth Monitoring Tools for Health Care Providers Managing Labor: A Scoping Review
JMIR Med Inform 2017;5(2):e14
URL: http://medinform.jmir.org/2017/2/e14/
doi:10.2196/medinform.6959
PMID:28619702

Original Paper

Effective Information Extraction Framework for Heterogeneous Clinical Reports Using Online Machine Learning and Controlled Vocabularies

Shuai Zheng[1], PhD; James J Lu[2*], PhD; Nima Ghasemzadeh[3], MD; Salim S Hayek[3], MD; Arshed A Quyyumi[3], MD; Fusheng Wang[4*], PhD

[1]Department of Biomedical Informatics, Emory University, Atlanta, GA, United States
[2]Department of Mathematics and Computer Science, Emory University, Atlanta, GA, United States
[3]Division of Cardiology, Emory School of Medicine, Emory University, Atlanta, GA, United States
[4]Department of Biomedical Informatics, Stony Brook University, Stony Brook, NY, United States
[*]these authors contributed equally

Corresponding Author:
Fusheng Wang, PhD
Department of Biomedical Informatics
Stony Brook University
2211 Computer Science
Stony Brook, NY, 11794-8330
United States
Phone: 1 6316327528
Fax: 1 6316381323
Email: fusheng.wang@stonybrook.edu

Abstract

Background: Extracting structured data from narrated medical reports is challenged by the complexity of heterogeneous structures and vocabularies and often requires significant manual effort. Traditional machine-based approaches lack the capability to take user feedbacks for improving the extraction algorithm in real time.

Objective: Our goal was to provide a generic information extraction framework that can support diverse clinical reports and enables a dynamic interaction between a human and a machine that produces highly accurate results.

Methods: A clinical information extraction system IDEAL-X has been built on top of online machine learning. It processes one document at a time, and user interactions are recorded as feedbacks to update the learning model in real time. The updated model is used to predict values for extraction in subsequent documents. Once prediction accuracy reaches a user-acceptable threshold, the remaining documents may be batch processed. A customizable controlled vocabulary may be used to support extraction.

Results: Three datasets were used for experiments based on report styles: 100 cardiac catheterization procedure reports, 100 coronary angiographic reports, and 100 integrated reports—each combines history and physical report, discharge summary, outpatient clinic notes, outpatient clinic letter, and inpatient discharge medication report. Data extraction was performed by 3 methods: online machine learning, controlled vocabularies, and a combination of these. The system delivers results with F1 scores greater than 95%.

Conclusions: IDEAL-X adopts a unique online machine learning–based approach combined with controlled vocabularies to support data extraction for clinical reports. The system can quickly learn and improve, thus it is highly adaptable.

(JMIR Med Inform 2017;5(2):e12) doi:10.2196/medinform.7235

KEYWORDS

information extraction; natural language processing; controlled vocabulary; electronic medical records

Introduction

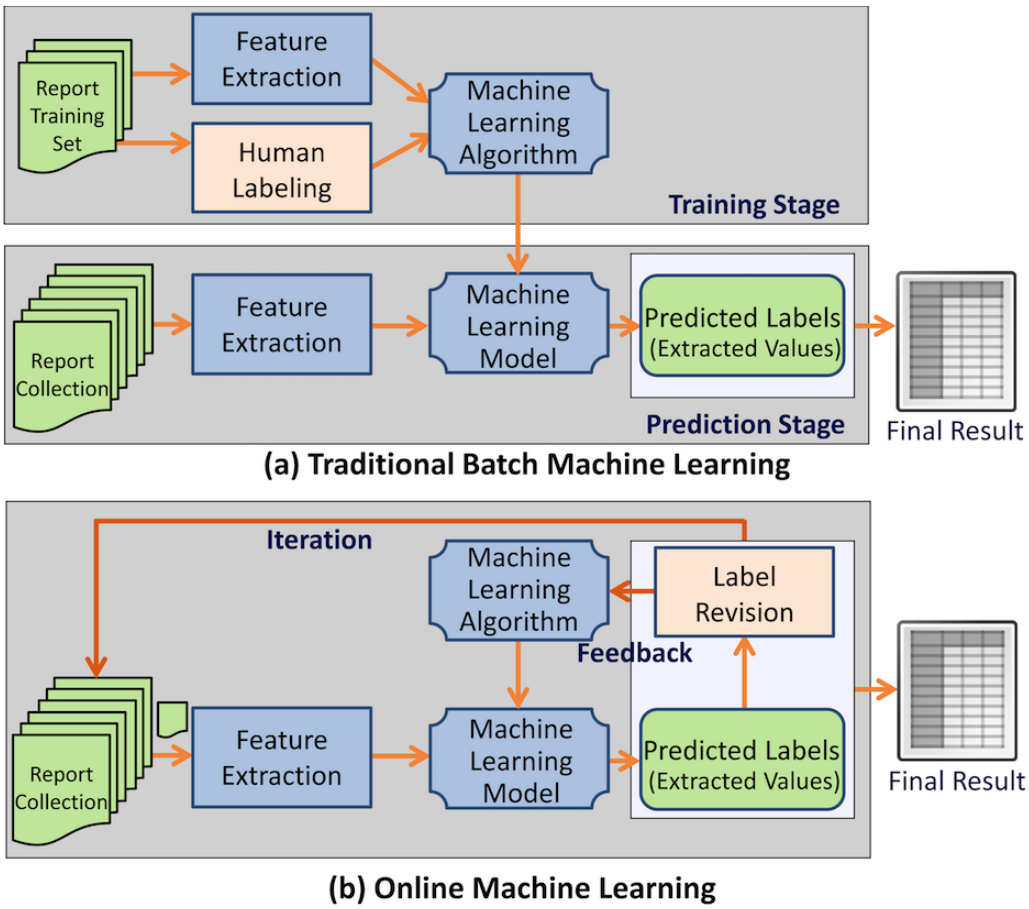
While immense efforts have been made to enable structured data model for electronic medical record (EMR), a large amount of medical data remain in free-form narrative text, and useful data from individual patients are usually distributed across multiple reports of heterogeneous structures and vocabularies. This poses major challenges to traditional information extraction systems, as either costly training datasets or manually crafted rules have to be prepared. These approaches also lack the capability of taking user feedbacks, to adapt and improve the extraction algorithm in real time.

Our goal is to provide a generic information extraction framework that is adaptable to diverse clinical reports, enables a dynamic interaction between a human and a machine, and produces highly accurate results with minimal human effort. We have developed a system, Information and Data Extraction using Adaptive Online Learning (IDEAL-X), to support adaptive information extraction from diverse clinical reports with heterogeneous structures and vocabularies. The system is built on top of online machine learning and customizable controlled vocabularies. A *demo video* can be found on YouTube [1].

IDEAL-X uses online machine learning–based approach [2-4] for information extraction. Traditional machine learning algorithms take a two-stage approach: batch training based on an annotated training dataset, and batch prediction for future datasets based on the model generated from stage one (Figure 1). In contrast, online machine learning algorithms [2,3] take an iterative approach (Figure 1). It learns one document at a time, and predicts values to be extracted for the next one. Learning occurs from revisions made by the user, and the updated model is applied to prediction for subsequent documents. Once the model achieves a satisfactory accuracy, the remaining documents may be processed in batch. Online machine learning not only significantly reduces human's effort for annotation but also provides the mechanism for collecting feedback from human-machine interaction to improve the system's model continuously.

Besides online machine learning, IDEAL-X allows for customizable controlled vocabularies to support data extraction from clinical reports, where a vocabulary enumerates the possible values that can be extracted for a given attribute. (The X in IDEAL-X represents the controlled vocabulary plug-in.) The use of online machine learning and controlled vocabularies is not mutually exclusive; they are complementary, which provide the user with a variety of modes for working with IDEAL-X.

Figure 1. Online machine learning versus batch learning. (a) Batch machine learning workflow; (b) Online machine learning workflow.



Background

Related Work

A number of research efforts have been made in different fields of medical information extraction. Successful systems include caTIES [5], MedEx [6], MedLEE [7], cTAKES [8], MetaMap [9], HITEx [10], and so on. These methods either take a rule-based approach, a traditional machine learning–based approach, or a combination of both.

Different online learning algorithms have been studied and developed for classification tasks [11], but their direct application to information extraction has not been studied. Especially in the clinical environment, the effectiveness of these algorithms is yet to be examined. Several pioneering projects have used learning processes that involve user interaction and certain elements of IDEAL-X. I^2E^2 is an early rule-based interactive information extraction system [12]. It is limited by its restriction to a predefined feature set. Amilcare [13,14] is adaptable to different domains. Each domain requires an initial training that can be retrained on the basis of the user's revision. Its algorithm $(LP)^2$ is able to generalize and induce symbolic rules. RapTAT [15] is most similar to IDEAL-X in its goals. It preannotates text interactively to accelerate the annotation process. It uses a multinominal naïve Baysian algorithm for classification but does not appear to use contextual information beyond previously found values in its search process. This may limit its ability to extract certain value types.

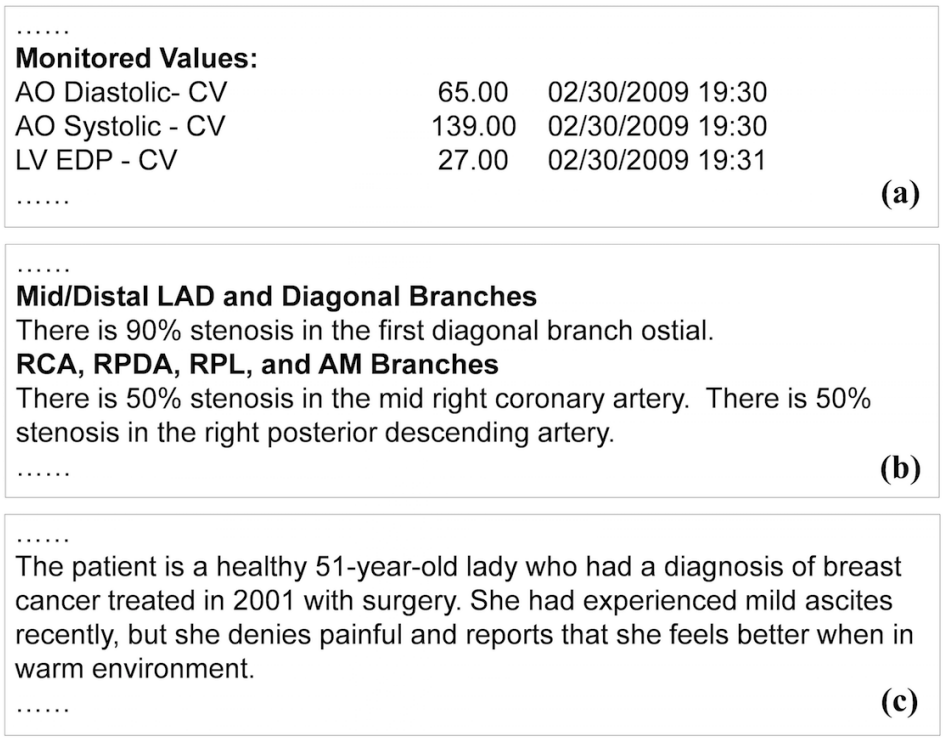
Different from online machine learning but related is active learning [16,17], it assumes the ability to retrieve labels for the most informative data points while involving the users in the annotation process. DUALIST [18] allows users to select system-populated rules for feature annotation to support information extraction. Other example applications in health care informatics include word sense disambiguation [19] and phenotyping [20]. Active learning usually requires comprehending the entire corpus in order to pick the most useful data point. However, in a clinical environment, data arrive in a steaming fashion over time that limits our ability to choose data points. Hence, an online learning approach is more suitable.

IDEAL-X adopts the Hidden Markov Model for its compatibility with online learning, and for its efficiency and scalability. We will also describe a broader set of contextual information used by the learning algorithm to facilitate extraction of values of all types.

Heterogeneous Clinical Reports

A patient's electronic medical record could come with a variety of medical reports. Data in these reports provide critical information that can be used to improve clinical diagnosis and support biomedical research. For example, the Emory University Cardiovascular Biobank [21] collects records of patients with potential or confirmed coronary artery diseases undergoing cardiac catheterization, and aims to combine extracted data elements from multiple reports to identity patients for research. Report types include history and physical report, discharge summary, outpatient clinic note, outpatient clinic letter, coronary angiogram report, cardiac catheterization procedure report, echocardiogram report, inpatient report, and discharge medication lists.

Figure 2. Example snippets of different report forms. (a) Semistructured report; (b) Template based narration; and (c) Complex narration.



We classify clinical reports into 3 forms: semistructured data, templatebased narration, and complex narration. Semistructured data represent data elements in the form of attribute and value pairs (Figure 2). Reports in this form have simple structures, making data extraction relatively straightforward. Template-based narration is a very common report form. The narrative style, including sentence patterns and vocabularies, follow consistent templates and expressions (Figure 2). Extracting information from this type of text (eg, "right posterior descending artery") require major linguistics expertise, to either formulate extraction rules or to annotate training data. Complex narration is essentially free-form text. It can be irregular, personal, and idiomatic (Figure 2). Most medical reporting systems still allow for (and thus encourage) such a style. It is the most difficult form to interpret and process by NLP algorithms. Nevertheless, certain type of information such as diseases and medications has finite vocabulary that could be used to support data extraction.

Methods

Overview

The interface and workflow conform to traditional annotation systems: a user browses an input document from the input document collection and fills out an output form. On loading each document, the system attempts to fill the output form automatically with its data extraction engine. Then, a user can review and revise incorrect answers. The system then updates its data extraction model automatically based on the user's feedbacks. Optionally, the user may provide a customized controlled vocabulary to further support data extraction and answer normalization. Pretraining with manually annotated data is not required, as the prediction model behind the data extraction engine can be established incrementally through online learning, customizing controlled vocabularies, or a combination of the two.

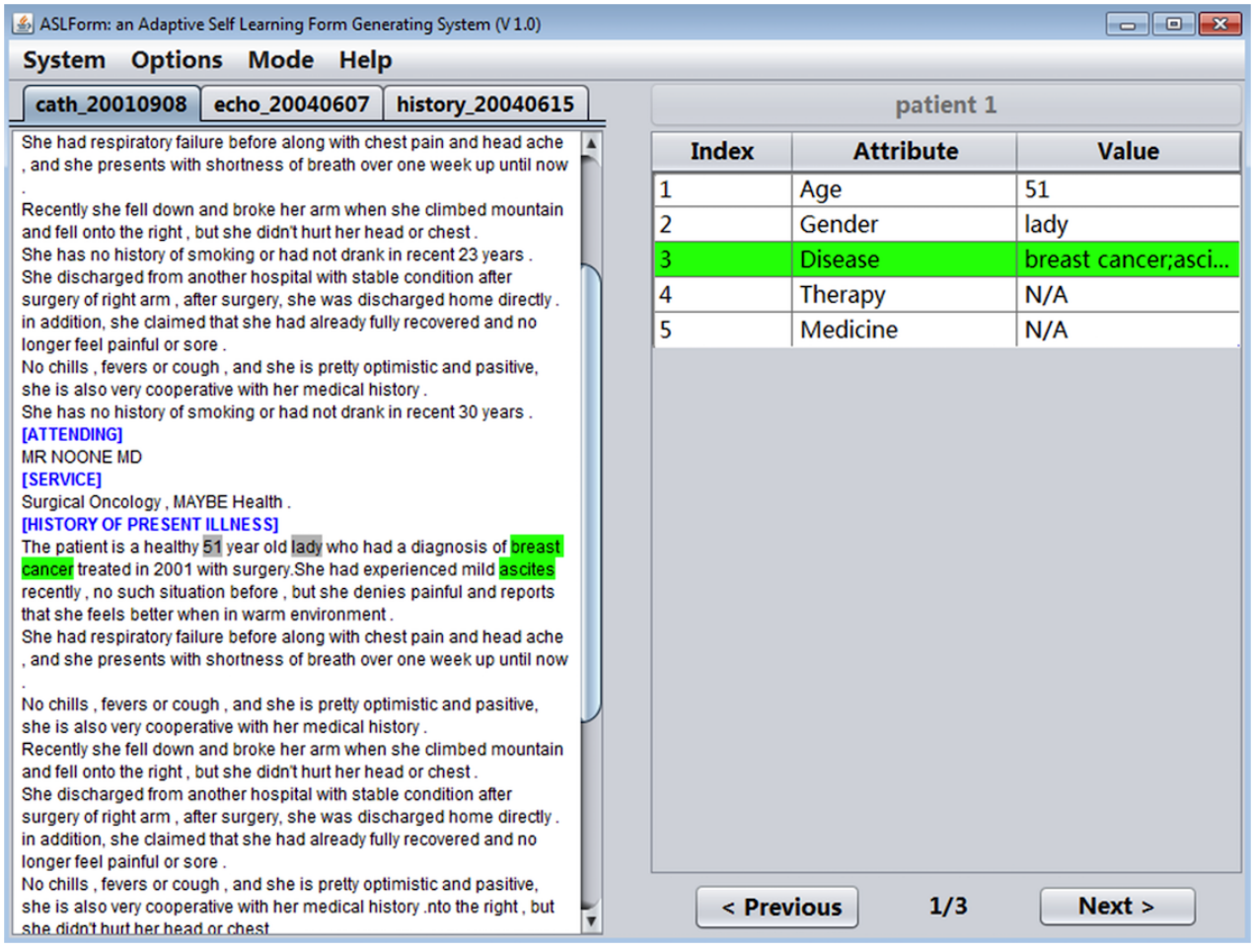
The system can operate in two modes: (1) interactive: through online learning, the system predicts values to be extracted for each report, and the user verifies or corrects the predicted values; and (2) batch: batch predicting for all unprocessed documents once the accrued accuracy is sufficient for users. Whereas interactive mode uses online machine learning to build the learning model incrementally, batch mode runs the same as the prediction phase of batch machine learning.

System Interface and User Operations

System Interface

IDEAL-X provides a GUI with two main panels: a menu and navigation buttons (Figure 3). The left panel is for browsing an input report, and the right panel is the output table with predicted values of each data element in the report. The menu provides options for defining the data elements to be extracted, specifying input reports, among others.

Figure 3. An example screenshot of IDEAL-X's interface.



Definition of Data Elements for Extraction

The system provides a wizard for constructing the metadata of the output form. The user builds the form by specifying a list of data elements and their constraints. An example is the data element "Heart Rate," which is constrained to be a numerical value between 0 and 200. Other constraints include sections of the report that may contain the values. However, except for the names of the data elements, specifying constraints are optional, as these can be learned by the system.

Data Extraction Workflow

The user will first select a collection of input reports to be extracted from a local folder. By default, the system runs in an interactive mode, and one report will be loaded at a time on the left display panel. The user can make manual annotations by highlighting the correct value in the report text. Clicking the corresponding data field in the table assigns the value to the data element. If the system has prefilled the field of a data element with a predicted value, the user can provide feedback by fixing incorrect values. As the user navigates to the next document, the system compares the prefilled and the final values for the most recently processed document. Values that are unchanged or filled in by users are taken as positive instances, and values that have been revised are taken as negative instances. Both instances are incorporated into the online learning algorithm to be used by the data extraction for subsequent documents. By iterating through this process, the amount of information that the system is able to correctly prefill grows over time. Note that manual revision in this context is different from traditional human labeling. It is only necessary if there is a wrong prediction, thus humans' effort can be significantly saved. Once the decision model reaches an acceptable level of accuracy, the user has the option to switch to batch mode to complete extraction for the remaining documents. If a patient has multiple reports, the text input panel displays each report with a separate tab. Data extracted from all the reports are aggregated in the output.

Customization of Controlled Vocabularies

IDEAL-X also provides an interface for the user to customize a controlled vocabulary that can be used by the system for data extraction. The controlled vocabulary contains both terminology and structural properties. The terminology includes lists of values and their normalization mappings. For example, Disease terminology includes "Diabetes Mellitus" with variations "DM" and "Diabetes." It also defines inductions. For example, taking "Insulin" or "Metformin" indicates having Diabetes Mellitus. Structural properties provide positive and negative contextual information for giving terms. For example, to extract medications taken by patients, the "Allergies" section is a negative context and medicine names in the section will be skipped. Structural properties may also contain disambiguation terms that may further improve the precision of extraction. A simple example is that "intolerant" is a negative indicator for identifying "statin" as "statin intolerant" refers to different a concept. Controlled vocabularies can be a powerful tool to support data extraction: it can be used to locate sentences and chunks of possible values, and to perform normalization for extracted values, discussed in the next section.

The Data Extraction Engine

While the user interacts with IDEAL-X interface, the data extraction engine works transparently in the background. The engine has 3 major components: answer prediction, learning, and the learning model that the online learning process continuously updates (Figure 4). The system combines statistical and machine learning–based approaches with controlled vocabularies for effective data extraction.

Document Preprocessing

When a report is loaded, the text is first parsed into an in-memory hierarchical tree consisting of 4 layers: section, paragraph, sentence, and token. Apache OpenNLP [22] is used to support the parsing with its Sentence Detector, Tokenizer, and Part-of-Speech Tagger. A reverse index of tokens is created to support efficient keywords-based search. The index is used to find locations (eg, sections, paragraphs, sentences, and phrases) of a token, as well as its properties such as part of speech and data type. For example, given the token "DM," the system can quickly identify the section (eg, "History") and the containing sentences. Such token search is frequently performed in answer prediction, and the in-memory index structures enable high efficiency for such operations.

Answer Prediction

Predicting the value of each data element involves the following steps: (1) *Identifying target sentences* that are likely to contain the answer; (2) *Identifying candidate chunks* in the sentences; (3) *Filtering the chunks* to generate candidate values; (4) *Ranking candidate values* to generate (raw) values; (5) *Normalizing values*; and (6) *Aggregating values* from multiple reports. The workflow is shown in Figure 5.

Figure 4. Overview of System Workflow.

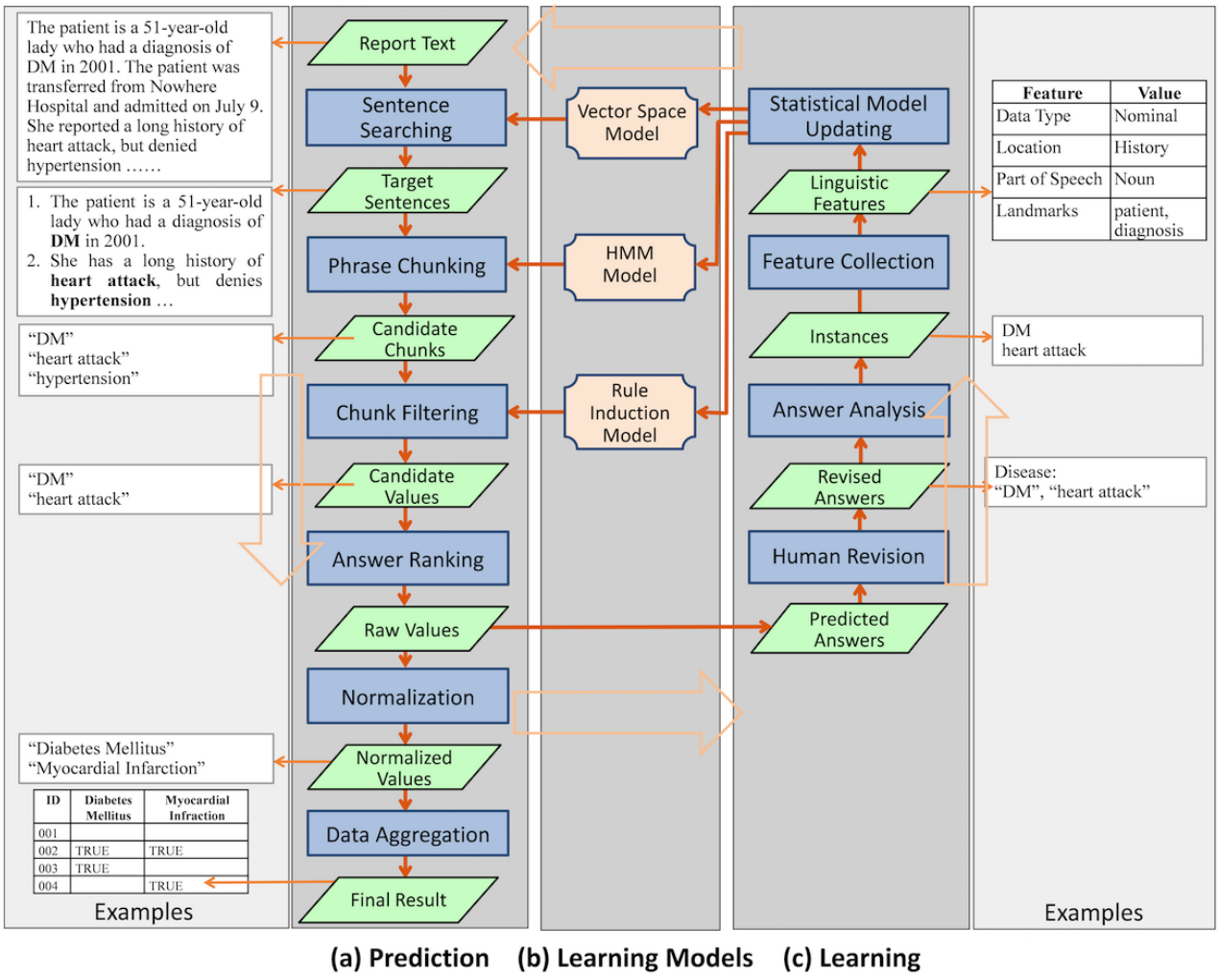
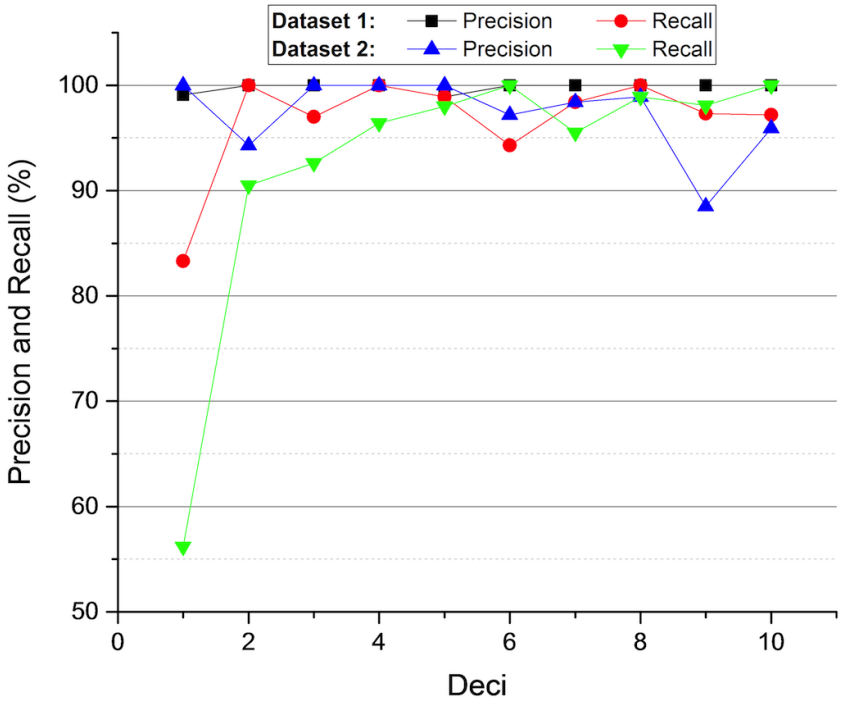


Figure 5. Precision and recall changes over processed records.



Identifying Target Sentences

Through online learning, the system accrues keywords from past answers (*answer keywords*) along with cooccurring words in the corresponding sentences (*contextual words*). For example, given the answer keywords "diabetes" and "hypertension" in the sentence "The patient reports history of diabetes and hypertension," contextual words are "patient," "report," and "history." Such answer keywords and contextual words combined with customized vocabularies can be utilized to identify sentences that are likely to contain answers with the following methods:

First, similarity-based search using the vector space model [23]. Given a collection of contextual words and their frequencies, the system computes the similarity against sentences in the document [23]. Sentences with high similarities are selected. For example, most sentences about "disease" contain "diagnosis" and "history." The past contextual keywords and their frequency weights are represented and maintained through a learning model discussed later in "Learning" section.

Second, answer keyword matching search. The answer keywords, combined with relevant user customized vocabularies, are also used to identify target sentences with keyword matching. For example, to extract diseases, if a sentence contains the disease term "myocardial infarction" defined in the vocabulary, the sentence is selected as a target. In both approaches, sections to be searched or skipped are also considered to narrow the scope of searching.

Identifying Candidate Chunks

After target sentences are selected, the system identifies potential phrases in the sentences using 2 methods: Hidden Markov model (HMM) [24] and keyword-based search. The HMM represents target words and contextual words in a sentence with different states, and marks values to be extracted based on probability distributions learned from previously collected values and their sentences. The keyword-based search finds candidate chunks using keywords collected from past answers and the controlled vocabulary.

Filtering chunks

To filter candidate chunks, the system uses rule induction [14,25] to generate "If-Then" rules based on historical statistics. The following filtering criteria are used: (1) Part of speech (POS): This filters a phrase by its POS tag in the sentence. Simple example phrases are noun and verb phrases. (2) String pattern: This looks for chunks that match special string patterns. For example, the first characters of all tokens are capitalized. (3) Value domain: This eliminates numerical or enumerated values that fall outside a specified range of values. (4) Negation: Based on predefined built-in rules, this removes phrases governed by words that reverse the meaning of the answer [26]. For example, if a candidate chunk "cancer" is extracted from a sentence "the patient has no history of cancer," "cancer" would not be included. (5) Certainty: Similar to negation filter, this detects and filters uncertain events or situations such as future plans, based on predefined rules. For example, a candidate chunk "radiation therapy" for treatment from a sentence "the patient is planned to take radiation therapy" should not be included. Whereas negation and certainty filtering is based on predefined rules, other filtering relies on real-time data statistics for filtering criteria.

Ranking Candidate Values

The system combines the scores of the selected sentences and chunks for ranking of candidate values. For a single-valued data element (eg, heart beat), the candidate value with the highest confidence score is selected. For a multi-valued data element (eg, medication), values with confidence scores above a threshold are selected. Based on this, each candidate value is either accepted or rejected.

Normalizing Values

This step normalizes extracted values through transformation, generalization, and induction rules given by the controlled vocabulary (Figure 4). For example, "DM" is transformed into "Diabetes Mellitus." "Pindolol" is generalized to its hypernym "beta blocker." The appearance of medication term "Metformin" (a drug for treating type 2 diabetes) in the text can infer the disease "Diabetes Mellitus."

Aggregating Results

Data extracted from multiple reports of a patient will be aggregated into a single table. The aggregation process may normalize values and remove duplicates. For example, "lisinopril" and "captopril" are extracted from discharge summary and inpatient report, respectively, and they can be normalized as "ACE inhibitor." If the same data element is extracted from multiple reports, deduplication is performed. The final output is in simple structural table form that can be exported conveniently to other applications such as Excel (Microsoft) or a database.

Note that controlled vocabularies can play important roles in the answer prediction process. They are used for identifying target sentences through keyword searching, identifying candidate chunks through keyword matching, and supporting normalization for extracted values.

Learning

IDEAL-X takes an online learning–based approach to incrementally build statistical models and make predictions (Figure 5). The 3 models used in IDEAL-X are all statistical based and can be continuously updated after each iteration.

System-predicted values automatically populate the output table, and the user advances to the next report with or without revision to these values. In both cases, the internal learning and prediction models of IDEAL-X are updated. For each instance, IDEAL-X collects and analyzes the following features: (1) Position: location of the answer in the text hierarchy; (2) Landmark: co-occurring contextual keywords in a sentence; (3) POS: parts of speech tag; (4) Value: the tokens of the answer; (5) String patterns: literal features such as capitalization and initial and special punctuation. These features are then used to update the 3 models.

In IDEAL-X, each data element such as attribute "disease" or "medicine," has its own statistical model, and each new instance of a data element will update the corresponding model. There

are 3 models to be updated: (1) Updating Space Vector Model: This model uses "Landmark" features of positive instances. The system updates frequencies of cooccurring contextual words, used as weights of the space vector [23]. (2) Updating HMM: HMM lists all words in a sentence as a sequence, in which an extracted value is marked as target value state and other words are recognized as irrelevant contextual states. Based on this sequence, the state transition probabilities and emission probabilities are recalculated [24]. (3) Updating rule induction model: Filtering rules are induced based on the coverage percentage [25]. Features such as POS, value domain and string patterns of both positive and negative instances are analyzed and their respective coverage percentages are modified. Once the coverage of a rule reaches a predefined threshold, the rule is triggered for filtering.

In an interactive mode, the above 4 steps repeat for each report, where the learning models are continuously updated and improved.

Results

Experimental Setup

Datasets

We used 3 datasets from 100 patients that were randomly sampled from a collection of about 5000 patients in the Emory Biobank database. Dataset 1 is a set of semistructured reports and contains 100 cardiac catheterization procedure reports. Dataset 2 is a set of template-based narration and contains 100 coronary angiographic reports. Dataset 3 is a set of complex narration and contains 315 reports, including history and physical report, discharge summary, outpatient clinic notes, outpatient clinic letter, and inpatient discharge medication report.

Ground Truth

The test datasets are independently hand-annotated by domain expert annotators, including physicians, physician trainees, and students trained by the Emory Clinical Cardiovascular Research Institute for Biobank data reporting. Each record is annotated by 2 different annotators. The interrater agreement scores (kappa) of these 3 datasets are .991, .986, and .835, respectively. An arbitrator—an independent cardiovascular disease researcher reconciles incompatible outputs of the system and the manual annotations to produce the final ground truth.

Evaluation Metrics

For validation, precision, recall, and F1 scores are used to estimate the effectiveness of extraction by comparing the system predicted results (before human revision) and the ground truth.

Experiment Settings

We aimed to evaluate the effectiveness of the system with respect to using online learning and controlled vocabularies and to understand their applicability to different report forms. By analyzing the report styles and vocabularies, we discovered that online learning will be more suitable for semistructured or template-based narration reports, and controlled vocabulary-guided data extraction would be more effective on complex narration with a finite vocabulary. Thus, we designed 3 experiments: (1) Online learning–based data extraction, where controlled vocabularies are not provided, based on Dataset 1 (semistructured) and Dataset 2 (template-based narration); (2) Controlled vocabularies-based data extraction, where online learning is not used, based on Dataset 3 (complex narration); and (3) Controlled vocabularies guided data extraction combined with online learning, based on Dataset 3.

Performance Evaluation

Experiment 1: Online machine Learning–Based Data Extraction

This experiment was based on Datasets 1 and 2. The system starts in an interactive mode with an empty decision model without prior training. The defined data elements are summarized in Multimedia Appendix 1. The user processes one report at a time, and each system-predicted value (including empty values for the first few reports) before user revision was recorded for calculating precision and recall.

Results are summarized in Table 1 for the 2 datasets, respectively. Both test cases achieved high precision as semistructured and template-based text is most easy to handle. To study the learning rate of online learning, we divided records into 10 groups, and plotted precision and recall of every 10% of the records in datasets 1 and 2. We observed that in both tests, the system maintained high precision during the learning process. Although some variability exists due to new data pattern, the recall of both cases also improved steadily. Not surprisingly, the rate of learning for dataset 1 is much faster given its semistructure.

Experiment 2: Controlled Vocabularies-Guided Data Extraction

In this experiment, online learning was disabled and data extraction was performed in batches using controlled vocabulary. Diseases and medications were extracted from Dataset 3 (values to be extracted are shown in Multimedia Appendix 1). Customized controlled vocabularies, including terminology and structural properties, had been created independently by physicians through referring to domain knowledge resources and analyzing another development report dataset of 100 patients, disjoint from Dataset 3. Note that comparisons in this and the following experiments were at a clinical finding level between system-integrated-results and manual-annotation-integrated-results.

Table 1. Results of data extraction from semistructured reports (Dataset 1) and template-based narration (Dataset 2).

Dataset	Numbers of data elements	Number of values	Precision (%)	Recall (%)	F1 (%)
1	19	1272	99.8	96.5	98.1
2	16	728	97.2	93.2	95.2

Table 2. Results of controlled vocabularies-guided data extraction from complex narration (Dataset 3).

Type of data elements	Number of data elements	Number of round truth values	Precision (%)	Recall (%)	F1 (%)
Diseases	15	418	94.5	99.0	96.7
Medications	10	437	98.6	99.7	99.2
All	25	855	96.5	99.4	97.9

The results in Table 2 show that controlled vocabularies are highly effective for data extraction over complex narratives. Domain-specific data, for example, cardiology-related diseases and medications, have limited numbers of possible values (or domain values), and a carefully customized controlled vocabulary can achieve high extraction accuracy.

Experiment 3: Controlled Vocabularies-Guided Data Extraction Combined With Online machine Learning

In this experiment, we performed 2 tests to examine how efficient and effective the system learns when only terminology is available and structural properties need to be obtained from online learning. Test 1 was to generate the baseline for comparison, and Test 2 was to demonstrate the effectiveness of combining online machine learning and controlled vocabularies. Dataset 3 was used to extract all diseases and medications.

For Test 1, terminology was used and online machine learning is disabled, so the test was guided by controlled vocabulary without any structural properties. We note that comprehensive terminology contributes directly to high recall rate, which means that the system seldom misses values to be extracted. However, if structural properties are not included, compared with the result in Experiment 2, the precision is much lower. This highlights the value of positive and negative contexts in an extraction task.

For Test 2, both terminology and online machine learning were used. Online machine learning supports learning structural properties. To show how quickly the system learns, only the 38 reports associated with the first 10 patients were processed with interactive online learning. All remaining reports were processed in batch. Results in Table 3 show an overall precision of 94.9%, which demonstrates that online learning could quickly learn structural properties.

Typical errors in these 2 tests were associated with terminology and contextual information used in complex narrative scenarios. On one hand, the completeness of the terminology list, including terms and their synonyms, influences the recall rate directly. On the other hand, although coverage of terminologies could be maximized by a carefully engineered vocabulary, unwanted extractions arising from searches in the wrong section, undetected negations, and ambiguous use of terms can still lower the overall precision.

Table 3. Results of controlled vocabularies-guided data extraction combined with online learning.

Test	Controlled vocabulary	Online learning	Precision (%)	Recall (%)	F1 (%)
1	Terminology only	N/A	80.9	99.4	89.2
2	Terminology only	Applied to first 10 patients	94.9	99.4	97.1

Discussion

Principal Findings

IDEAL-X provides a generic data extraction framework that takes advantage of both online learning and controlled vocabularies. The 2 approaches complement each other and can also be combined. Online learning–based approach is highly effective for reports with underlying structural patterns such as semistructured or template-based narration style-based reports. Experiments with complex narrative reports indicate that the use of controlled vocabularies is highly effective for supporting extraction constrained by finite data domain. In addition, structural properties such as section-data associations can play an important role in improving the accuracy of extraction. However, in cases where controlled vocabularies are unavailable—extracting generic named entities for example, maintaining high accuracy is a challenge. This is an ongoing area of exploration that we will report in the future.

Machine learning is among major techniques for identifying candidate chunks. Besides HMM, we have also explored other classifiers such as Naive Bayes classifier and neural networks-based classifier. An ongoing project includes a systematic study of different classifiers and their combinations (including Conditional Random Field and Support Vector Machine [27]) for online machine learning–based data extraction.

To make it more flexible on using standard medical terminologies for customizing controlled vocabularies, an ongoing work is developing a tool that can easily search and import concepts from standard vocabularies such as ICDE-9, ICD-10, and SNOMED, from a local file or through NCBO BioPortal.

Conclusions

Although there are natural language processing tools available for extracting information from clinical reports, the majority lack the capability to support interactive feedback from human users. An interactive, online approach allows the user to coach the system using knowledge specific to the given set of reports, which may include local reporting conventions and structures. Moreover, no advanced linguistics knowledge or programming skills are required of the users; the system maintains the ordinary workflow of manual annotation systems. We perform a systematic study on the effectiveness of the online learning–based method combining with controlled vocabularies

for data extraction from reports with various structural patterns, and conclude that our method is highly effective. The framework is generic and the applicability is demonstrated with diverse report types. The software will be made freely available online [28].

Acknowledgments

The study was funded by Center for Disease Control and Prevention, contract # 200-2014-M-59415.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Test cases.

[PDF File (Adobe PDF File), 32KB - medinform_v5i2e12_app1.pdf]

References

1. Youtube. IDEAL-X Demo Video URL: https://www.youtube.com/watch?v=Q-DrWi31nv0&feature=youtu.be [accessed 2017-04-20] [WebCite Cache ID 6prjtKpLd]
2. Shalev-Shwartz S. Uchicago. 2007. Online learning: theory, algorithms, and applications URL: http://ttic.uchicago.edu/~shai/papers/ShalevThesis07.pdf [accessed 2017-05-03] [WebCite Cache ID 6qAyaWMUk]
3. Shalev-Shwartz S. Online learning and online convex optimization. FNT in Machine Learning 2011;4(2):107-194. [doi: 10.1561/2200000018]
4. Smale S, Yao Y. Online learning algorithms. Found Comput Math 2006;6(2):145-170. [doi: 10.1007/s10208-004-0160-z]
5. Crowley RS, Castine M, Mitchell K, Chavan G, McSherry T, Feldman M. caTIES: a grid based system for coding and retrieval of surgical pathology reports and tissue specimens in support of translational research. J Am Med Inform Assoc 2010;17(3):253-264 [FREE Full text] [doi: 10.1136/jamia.2009.002295] [Medline: 20442142]
6. Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: a medication information extraction system for clinical narratives. J Am Med Inform Assoc 2010;17(1):19-24 [FREE Full text] [doi: 10.1197/jamia.M3378] [Medline: 20064797]
7. Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. J Am Med Inform Assoc 2004;11(5):392-402 [FREE Full text] [doi: 10.1197/jamia.M1552] [Medline: 15187068]
8. Savova G, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17(5):507-513 [FREE Full text] [doi: 10.1136/jamia.2009.001560] [Medline: 20819853]
9. Aronson A. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp 2001:17-21. [Medline: 11825149]
10. Zeng Q, Goryachev S, Weiss S, Sordo M, Murphy SN, Lazarus R. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. BMC Med Inform Decis Mak 2006 Jul 26;6:30 [FREE Full text] [doi: 10.1186/1472-6947-6-30] [Medline: 16872495]
11. Hoi S, Wang J, Zhao P. Libol: a library for online learning algorithms. J Mach Learn Res 2014;15(1):495-499.
12. Cardie C, Pierce D. Cornell University. 1998. Proposal for an interactive environment for information extraction URL: http://www.cs.cornell.edu/home/pierce/papers/cucstr98.pdf [accessed 2017-05-03] [WebCite Cache ID 6qAzn2nLD]
13. Ciravegna F, Wilks Y. Designing adaptive information extraction for the semantic web in amilcare. In: Annotation for the semantic web. The Netherlands: IOS Press; 2003:96.
14. Ciravegna F. Adaptive information extraction from text by rule induction generalisation. 2001 Presented at: International Joint Conference on Artificial Intelligence; 2001; Seattle, Washington, DC.
15. Gobbel G, Garvin J, Reeves R, Cronin RM, Heavirland J, Williams J, et al. Assisted annotation of medical free text using RapTAT. J Am Med Inform Assoc 2014;21(5):833-841 [FREE Full text] [doi: 10.1136/amiajnl-2013-002255] [Medline: 24431336]
16. Settles B. WISC. 2010. Active learning literature survey URL: https://research.cs.wisc.edu/techreports/2009/TR1648.pdf [accessed 2017-05-03] [WebCite Cache ID 6qB0Le2O5]
17. Olsson F. Swedish-ICT. 2009. A literature survey of active machine learning in the context of natural language processing URL: http://soda.swedish-ict.se/3600/1/SICS-T--2009-06--SE.pdf [accessed 2017-05-03] [WebCite Cache ID 6qB0USdQo]
18. Settles B. Closing the loop: fast, interactive semi-supervised annotation with queries on features and instances. 2011 Presented at: Conference on empirical methods in natural language processing; 2011; Edinburgh, United Kingdom p. 1467-1478.

19. Chen Y, Cao H, Mei Q, Zheng K, Xu H. Applying active learning to supervised word sense disambiguation in MEDLINE. J Am Med Inform Assoc 2013;20(5):1001-1006 [FREE Full text] [doi: 10.1136/amiajnl-2012-001244] [Medline: 23364851]
20. Chen Y, Carroll RJ, Hinz ER, Shah A, Eyler AE, Denny JC, et al. Applying active learning to high-throughput phenotyping algorithms for electronic health records data. J Am Med Inform Assoc 2013 Dec;20(e2):e253-e259 [FREE Full text] [doi: 10.1136/amiajnl-2013-001945] [Medline: 23851443]
21. Eapen DJ, Manocha P, Patel RS, Hammadah M, Veledar E, Wassel C, et al. Aggregate risk score based on markers of inflammation, cell stress, and coagulation is an independent predictor of adverse cardiovascular outcomes. J Am Coll Cardiol 2013 Jul 23;62(4):329-337 [FREE Full text] [doi: 10.1016/j.jacc.2013.03.072] [Medline: 23665099]
22. Apache. Apache OpenNLP URL: http://opennlp.apache.org/ [accessed 2017-04-20] [WebCite Cache ID 6prnK27wY]
23. Manning CD, Raghavan P, Schütze H. Introduction to Information Retrieval. New York: Cambridge University Press; 2008.
24. Elliott RJ, Aggoun L, Moore JB. Hidden Markov models: estimation and control. New York: Springer Science & Business Media; 2008.
25. Fürnkranz J. Separate-and-conquer rule learning. Artificial Intelligence Review 1999;13(1):3-54.
26. Huang Y, Lowe HJ. A novel hybrid approach to automated negation detection in clinical radiology reports. J Am Med Inform Assoc 2007;14(3):304-311 [FREE Full text] [doi: 10.1197/jamia.M2284] [Medline: 17329723]
27. Jiang M, Chen Y, Liu M, Rosenbloom ST, Mani S, Denny JC, et al. A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries. J Am Med Inform Assoc 2011;18(5):601-606 [FREE Full text] [doi: 10.1136/amiajnl-2011-000163] [Medline: 21508414]
28. IDEAL-X. URL: http://www.idealx.net/home.html [accessed 2017-05-04] [WebCite Cache ID 6qCYtYNoU]

Abbreviations

EMR: electronic medical record
HMM: hidden markov model
NLP: natural language processing
POS: part of speech

Edited by G Eysenbach; submitted 28.12.16; peer-reviewed by M Jiang, NP Cruz-Díaz; comments to author 09.02.17; revised version received 16.03.17; accepted 20.03.17; published 09.05.17.

Please cite as:
Zheng S, Lu JJ, Ghasemzadeh N, Hayek SS, Quyyumi AA, Wang F
Effective Information Extraction Framework for Heterogeneous Clinical Reports Using Online Machine Learning and Controlled Vocabularies
JMIR Med Inform 2017;5(2):e12
URL: http://medinform.jmir.org/2017/2/e12/
doi:10.2196/medinform.7235
PMID:28487265

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org