RESEARCH ARTICLE

Developing a Standardized Medical Speech Recognition Database for Reconstructive Hand Surgery

Fateme Sadeghi [1*] , Marjan Ghazisaeedi [2], Reza Safdari [2], Abdoljalil Kalantarhormozi [3]

[1] Department of Health Information Technology, School of Allied Health Sciences, Tehran University of Medical Sciences, Tehran, Iran [2] Department of Health Information Management, School of Allied medicine, Tehran University of Medical Sciences, Tehran, Iran [3] School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Abstract

Fast and holistic access to the patients' clinical record is a major requirement of modern medical Decision Support Systems (DSS). While Electronic Health Records (EHRs) have replaced the traditional paper-based records in most healthcare organizations, the data entry into these systems remains largely manual. Speech recognition technology promises substitution of the more convenient speech-based data entry with currently laborious manual method in the near future. Developing effective Speech Recognition Systems (SRS) requires availability of standardized vocabulary databases. This study aimed at developing a medical speech recognition database for reconstructive hand surgery based on the Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT). All codes related to hand problems were extracted from ICD-10. A sample of 2051 diagnosis codes was randomly selected from the patients' records. The operation report paper sheets were transformed into electronic records. For each term, the SNOMED-CT was browsed to find the preferred synonym using CliniClue® Xplore Software. For some words with several number of synonyms, the preference of reconstructive surgery specialists was asked using a researcher-made questionnaire. Ultimately, the preferred words were substituted throughout each document and used for developing a database of standard nomenclature. The developed database was used in speech-based recording clinical data in reconstructive hand surgery operating room, and an accuracy of 81% in correct recording of clinical data was observed. Therefore, development of standard medical nomenclature databases can facilitate accurate electronic recording of medical data, and reduce the associated labor and cost posed by the current manual method.

Keywords: Clinical records, Health Electronic Systems, Electronic Health Records, Speech recognition, Database, Software, Medical Informatics, Hand surgery

Background and Objectives

Application of ICT in health (eHealth) has become an integral part of modern healthcare systems [1, 2]. Converting paper records to into EHR (Electronic Health Record) is an important project in all healthcare organizations [3]. The most important option for health information exchange is to organize large volumes of patient data at the time of creation [4]. Electronic medical records need consistent and valid data elements. However, the healthcare centers use several nomenclatures for classifying data elements. Accordingly, it is difficult to integrate and exchange information[3]. The use of standard terminology and a unified classification system for naming and coding the data elements will enhance the consistency and validity of health information. Thus, the collected data can be used for analyzing the patient data and quality monitoring, determining the optimal performance, and supporting clinical and practical research [5]. Therefore, a common clinical vocabulary will be necessary to be used in the electronic systems as a basis for communication and correct data reporting [3].

However, there is not yet a comprehensive dictionary to meet the all requirements of information exchange in the healthcare sector [4]. Medical nomenclature systems are necessary for recording, retrieval and exchange of clinical data for the electron-

Corresponding author: Fateme Sadeghi, Health Information Technology, School of Allied Health Sciences, Tehran University of Medical Sciences, Tehran, Iran, Tel: +98 9125900612, Email: Fateme_S2860@yahoo.com

ic health records. The most comprehensive clinical reference terminology is the international systemized nomenclature of medicine-clinical terms (SNOMED-CT). SNOMED-CT is a common language to integrate the clinical data exchange of service providers leading to effective healthcare service [3, 4]. In this case, data recorded using SNOMED-CT provides a common language and integrated approach to record and collect health data for electronic health records from the healthcare professions and locations [3, 6].

To create a computer record, the patient data should be typed by clinical staff. Despite the objections of the clinical staff towards the use of information systems, most of them are challenged for data entry into computer. Usually, the time needed to record data is longer than that needed for dictation or handwrite [7]. A documentation approach to complete the patients' records without the use of computer data will promote the data exchange [8]. Speech recognition software (SRS) is of new technologies, which can be used in this regard. It will facilitate the data entry into the patients' records. SRS defines relevant database in the internal software. Therefore, a standardized database will give rise to a unified standard record throughout Iran [9]. Thus, standardization of data collection tools is considered as a part of standardization of health records. Accordingly, the design, implementation, monitoring, and review of data collection tools play an important role in improving the quality of data [10]. The aim of the present study is to develop a preferred speech recognition database in the field of reconstructive hand surgery using SNOMED-CT. It will provide an integrated standard tool to facilitate data entry and exchange of patients' records.

Methods

This is a quasi-experimental intervention study conducted in 15 Khordad Hospital affiliated to Shahid Beheshti University of Medical Sciences (Tehran, Iran) from March 20 to June 16, 2012 using a cross-sectional approach. First, all codes related to hand problems were extracted from ICD-10. Then the number of patients referred to 15 Khordad Hospital in 2011 was extracted in terms of diagnostic code. The total number of diagnostic codes was equal to 16.053. After consultation with statistical experts, the sample was selected using Table 1 considering

at least 15 referred cases for each diagnostic code. The population consisted of all records of inpatients with hand surgery (627 records with 2051 diagnostic codes in ICD-10).

In the case where the number of diagnostic codes was less than 15, the records of patients in the previous year (2011) were used to reach 15 codes. The operation report sheets were then converted into an electronic format, and the words in the electronic files were preferred. After finding each preferred word, it was replaced in the entire text. A research-made questionnaire was also used to select some of the preferred words. The content validity of the questionnaire was assessed by the comments of the faculty of the Department of Health Information Technology, Tehran University of Medical Sciences (Tehran, Iran) and the faculty of the Department of Plastic and Reconstructive Surgery, Shahid Beheshti University of Medical Sciences (Tehran, Iran). The reliability of the questionnaire was assessed by Test-Retest method. At this stage, the questionnaire was sent to 5 hand surgery specialists and 3 health information technology experts. They were asked to complete the questionnaire. After a week, the questionnaire was again completed by the same specialists. The results were compared with those obtained in the previous step. A Spearman's correlation coefficient of R=0.89 was obtained. Then the comments by 20 plastic and reconstructive surgery specialists were used to select the appropriate term, as well as several terms used in the operation report sheets among the SNOMED-CT terms. The prefabrication results obtained from the questionnaire were added to the previous sections.

Results

Each concept has a very specific name and a preferred term used by therapists to designate that concept [11]. Concepts may have one or more synonyms [6]. In this research, the preferred terms are those designated by Preferred in ClinClue Software and the electronic format of SNOMED-CT. According to the operation report sheets, because of the variety of words and terms, all words used in the operation report sheets were divided into five categories as shown in Table 2.

The various types of words were determined. According to the results, general Persian words showed

the highest frequency of 32%. Therefore, they remained unchanged in the database. English medical terms showed a repetition frequency of 29% in the texts, and were extended to the prefabrication section. The words in operation report sheets were generalized using the preferred words in SNOMED-CT. For this purpose, the selected word was defined as the preferred word, and other synonymous were defined as pronunciation of the preferred word. In transliteration section, reference was made to the preferred word. In fact, the words were not removed from the system; rather they were referred to the selected word. Therefore, once each word was expressed by the physician, only the preferred word in the database was typed in the operation report sheet. Figure 1 shows the prefabrication process.

According to the results presented in Table 2, the medical abbreviations in the operation report sheets include 17% of words in the text. The abbreviations were also preferred using SNOMED-CT. The medical terms and corresponding abbreviations were also provided. The Persian and English abbreviations in the operation report sheets were defined as single terms. In this section, the variety of words decreased from 82 to 17 types, leading to a uniform writing form in the whole text.

Prefabrication of all medical terms by SNOMED-CT caused disorganization of the sentences in the Persian texts. Therefore, the comments by specialists were used to select at least one term from a variety of Persian words with several writing forms (10% of English terminology and 12% of Persian medical terminology). In this section, the selected word was evaluated by 20 physicians using a questionnaire. Other terms were preferred by the physicians. In addition to a variety of words used in the operation report sheets, the preferred term in SNOMED-CT was also indicated for the physician. After collecting and analyzing the results of questionnaires, the words were first extracted in terms of pronunciation in SNOMED-CT, and then based on the physicians' comments, the extracted words were replaced in the original file.

After replacement of the selected words in the whole text, the pronunciation and language models were created in the database. Then the as-designed database was loaded in NEVISA to obtain the medical version of the software. The software was tested in the operating room of 15 Khordad Hospital. The results showed an approximate accuracy of 81%.

Table 1 The diagnostic codes related to prefabrication of operation report sheets

Number of records	Number of codes in the records
15 cases	< 50 codes
50% of referrals	50-100 codes
20% of referrals	100-200 codes
10% of referrals	200-500 codes
5% of referrals	500 codes <

Discussion

Clinical data standards play an important role in improving the quality and efficiency of healthcare [12, 13]. Without data standards, organizations will not be able to exchange clinical information between their systems [13]. To design a system, it is necessary to determine its requirements. If the present system is established based on the existing standards, the system will have a higher efficiency in future. Thus, you will not have to re-work and change the as-designed items to exchange information [10].

Table 2 The variety of words in the electronic texts in terms of the types of words

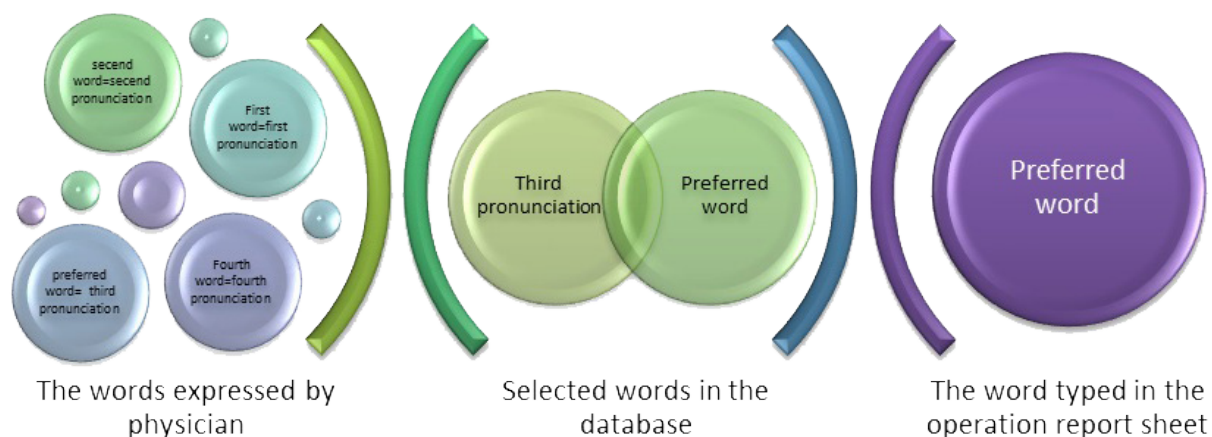|	Number of words	Diversity	
		Before prefabrication	After prefabrication
Pronunciation of English terms	3776	321	95
Persian medical terms	4450	368	112
English medical terms	10837	472	440
General Persian words	12096	652	652
Medical abbreviations	6466	82	25
Total	37625	1895	1324

Figure 1   The citation of medical terms expressed by the physician to a preferred term in the database to be typed in the operation report sheet.

The main objective of the present study is to create a system with a common basic standard. This is in line with previous studies.

Based on the results of relevant studies, vocabulary and terminology standards are essential to facilitate the exchange of information in electronic systems [14]. Also systems following different standards are unable to effectively communicate to exchange information. Hence, standardization and development of standards for integration of health information systems are considered as the most important issues in the management of relevant systems [15]. The results of the present study are consistent with those obtained by Amin-pour [15].

According to Safdari, before the development of electronic medical records, integrated standards should be developed in the field of structure, content, method of data exchange, and security of information systems. The information should be made initially in a standard form to be used by specialists during the data exchange. Also duplication costs to create a standard in later stages should be prevented [16]. These were considered in this study using the international SNOMED-CT terminology as the basic standard. This terminology provides a readable method for clinical information in the IT system. The main objective of the present study is to create an integrated

standard database for data entry into the patients' records in Iran. According to the results, the primary objective has been partly realized. This is in line with the findings of Donnelly [17].

The international SNOMED-CT system can be used to standardize clinical data in the patients' records to eliminate some design and implementation limitations of the EHR system. The use of data collection technologies is a practical solution to implement electronic medical records without affecting relationship between the healthcare providers and patients. For this purpose, SRS and voice input by microphone are suggested instead of writing a document. In this study, the two methods proposed by "Amatayakul" were employed [18].

Using this system in the electronic records, the care providers and patients will be able to share data at any time and place. This feature and centralized access to various concepts such as allergies to medicines, measures, diseases and harmful effects of drugs enable us to avoid many medical errors due to paper records. Finally, data should be effectively collected once and shared for different uses [19]. This is also consistent with the results of the present study.

With the advent of public healthcare, care organizations are needed to be integrated with the network

of healthcare providers to focus on the healthcare services. Therefore, it is essential to use appropriate healthcare standards for integrating the structure and terminologies in heterogeneous systems to provide appropriate technical infrastructure and to access integrated patient information exchange. This requirement can be achieved by using standard terminology systems such as ICD and/or SNOMED for programming patients' records [17].

One of the main goals of electronic records is to establish systems with mutual interaction capabilities. Obviously, systems following different standards are unable to communicate and coordinate with each other, and cannot improve patient care process. Hence, standardization and development of standards for integration of health information systems are of the main issues in the management of such systems [15].

Conclusions

The data obtained from the prefabrication possess a basic standard with an international infrastructure. In addition to integrated national vocabulary, it can be used to create thesaurus, different medical anthologies, and integrated medical texts, as well as national nomenclature systems. In addition, it facilitates data entry into HER while facilitating data exchange. The integration of software with Hospital Information System (HIS) will strengthen the software in terms of demographic data. Thus, it is a basis to create a faster and more accurate operation report sheet in the electronic records. The high accuracy of the system indicates that this standardized tool can be used in the healthcare network. The practical software will help the development of an integrated Electronic Health Record (EHR) system.

References

1. Mohammadzadeh N, Safdari R. Artificial intelligence tools in health information management. *Int J Hosp Res* 2012, 1(1):71-6.

2. Tabibi ST, Zaki TS, Ataeepoor Y. Developing an expert system for diabetics treatment advices. *Int J Hosp Res* 2013, 2(3):155-62.

3. Safdari R, Ghazisaeedi M, Partovipoor E. Comparative study of evolution process and structure of systematized nomenclature of medicine in America, England and Australia countries. *Payavard Salamat* 2008, 2(1):18-30.

4. Partovipoor E. Comparative study of systematic nomenclature of medicine in USA, England and Australia in 2006. *MS Thesis*. Tehran University of Medical Sciences, 2006.

5. Sadughi F, Samadbeik M, Ehteshami A, Aminpur F, Rezaee Hachesu P. *Health Information technology*. Tehran: Jafari; 2011.

6. Safdari R, Meidani Z. Health Services Accreditation Standards for information management in Canada, New Zealand and USA: a comparative study. *J Qazvin Uni Med Sci* 2007, 11(1):73-8.

7. Sadoughi F, Ghazisaeidi M, Meraji M, Kimiafar K, Ramezanghorbani N. *Health information management technology*. Tehran: Jafari 2011.

8. Yen PY, Gorman P. Usability testing of digital pen and paper system in nursing documentation. *AMIA Annu Symp Proc* 2005:844–8.

9. TSCo. Analysis of feasibility speech database produced in Farsi speech recognition engine. tehran: University of Science and Technology 2009.

10. Safdari R, Masuri N. *Electronic Health Record System (Standards, Health data)*. Tehran: Jafari; 2011.

11. Review of classification systems in the health sector. Tehran: Tehran University of Medical Sciences 2011.

12. Hosseinpourfard M, Abbasi S, Ayoubian A, Izadi M, Mahjoob MP. Hospital compliance with clinical documentation standards: A descriptive study in two Iranian teaching hospitals. *Int J Hosp Res* 2012, 1(2):121-5.

13. CHF. Clinical data standards explained. FactSheet. California: California Healthcare Foundation 2004.

14. Moghaddasi H, Hosseini A, Asadi F, Seyyedi SS. Evaluation of information system for organ transplantation in hospitals affiliated to medical universities in Tehran. *Pajoohandeh J* 2012, 16(7):333-9.

15. Sadoughi F, Aminpour F. How to standardize electronic medical records. *Ir J Med Educ* 2012, 11(9):1140-8.

16. Sadari R, Masoori N, Farajollah SSS. A comparative study on the functions of pioneer organizations (ASTM, HL7 and ISO) in developing electronic health record. *Health Inform Manage* 2011, 8(3):422-32. [Persian]

17. Bos L, Donnelly K. SNOMED CT: The advanced terminology and coding system for eHealth. *Stud Health Technol Inform* 2006, 121:279-90.

18. Amatayakul M. *Electronic health records: A practical guide for professionals and organizations.* Chicago: American Health Information Management Association; 2007.

19. Ahmadi M, Rezaei H, Shahmoradi L. *Electronic health record: structure, content, and evaluation.* Tehran: Jafari; 2008.

Please cite this article as:
Fateme Sadeghi , Marjan Ghazisaeedi, Reza Safdari, Abdoljalil Kalantarhormozi. Developing a Standardized Medical Speech Recognition Database for Reconstructive Hand Surgery. *International Journal of Hospital Research* 2014, **3**(3):149-154.