A Three-Phase Decision Model of Computer-Aided Coding for the Iranian Classification of Health Interventions (IRCHI)

Zahra Azadmanjir1, Reza Safdari*, Marjan Ghazisaeedi1, Mehrshad Mokhtaran1, Mohammad Esmail Kameli2

1Department of Health Information Management, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran
2Statistics and Treatment Information Management Unit, Deputy of Treatment, Ministry of Health and Medical Education of Iran, Tehran, Iran

Corresponding author: Reza Safdari, School of Allied Medical Sciences, Farredanesh Alley, Ghods Ave, Enghelab Ave, No #17, Tehran, Iran. Phone: +98 21 88983025; Fax: +98 21 88983037. E-mail: azadm@razi.tums.ac.ir.


ORIGINAL PAPER

INTRODUCTION

Data and statistics about medical procedures or other services that is delivered to health improvement, are critical from several aspects. It is not limited to surgeries or therapeutic procedures, but is related to all health interventions. In general, health interventions refers to any activity that is done to assess, promote or improve the health status and health-related functions of a patient or population of individuals (1). Data of interventions is important to following applications: The first, to measure the utilization of health services according to utilization indicators (2). The second, evaluating the quality of care in all of the healthcare settings such as acute care, rehabilitation and public health. Third, research on effectiveness or other effects of interventions and procedures on patients or specific populations. In addition, data of interventions can help to management and allocation of health resources through identifying the required interventions patterns based on population characteristics include age, gender and prevalent health problems in different regions (3).

The most important concern to use health intervention data in the aforementioned applications is the accuracy of these data, which are usually in the form of coded data by classification systems (4). In fact, the quality of coding of medical procedures or health intervention data will have a significant impact on the accuracy of the data that are used in the above fields. Due to the importance of this issue, many studies determine the accuracy and consistency of health coded data and solutions to improve. For example, A British systematic review showed that the median

© 2017 Zahra Azadmanjir, Reza Safdari, Marjan Ghazisaeedi, Mehrshad Mokhtaran, Mohammad Esmail Kameli

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
accuracy of diagnostic codes was 80.3% and for procedure codes was 84.2% (5). Similar systematic review in 2015 on accuracy of clinical coding in Iran, has been shown that the accuracy of diagnostic codes was between 41.8% to 88.78% and for procedure codes was between 80.21% to 98.93% (6).

Another study in the USA showed that eight percent of monthly coded imaging procedures related to the musculoskeletal system were wrong in hospitals (7). In the similar prospective study, error rate of assigned codes to the oral and maxillofacial surgeries was between 32 to 33% in teaching hospitals of England (8). The results of the similar study in Iran showed that from 246 medical records, 18.7% have an error in procedure codes (9).

It is said that, one of best solutions other than coders training and physician involvement in the coding work, is Computer-Assisted Coding (CAC) (10, 11). According to definition provided by American Health Information Management Association (AHIMA), CAC is means “the use of computer software that automatically generates a set of medical codes for review, validation and use based upon clinical documentation provided by healthcare practitioners” (12). There are different types of CAC solutions from fully automated by Natural Processing Language (NLP) techniques to semi-automated types. Selection of a proper method is not so easy and depends on the terms of use.

In Iran, from one side, with the implementation of the health reform plan from 2013, the quality of clinical documentation and medical coding were taken into account more than the past. Investigations in 2016 about the status of coding software in hospitals depicted that there is no specific system for coding and coders use the subsystem of Hospital Information System only for entering the codes and limited reporting. They do not have any the capability to assist the decision making of coders (13). Then, CAC systems are required.

From the other side, Iran is among the countries that have no national classification for procedure coding and use volume three of ICD-9-CM for it. Moreover, only invasive surgical and nonsurgical interventions are being coded based on ICD-9-CM volume 3 and other procedures only are being coded by Relative Value Units (RVU) coding system for billing and reimbursement. For this reason, efforts to develop a national classification have been started in October 2015, following the end of maintaining ICD-9-CM and replacing ICD-10-CM, PCS in United State. According to the recommendations of the World Health Organization to countries that have used the ICD-9-CM for procedure coding, were initiated activities and programs for developing a national adopted classification based on framework of International Classification of Health Interventions (ICH) (14). Now, development of CAC based on ICD-9-CM does not have economic justification. In contrast, CAC becomes indispensable when the new classification system will be replaced. Since, there will genuine concerns for future successful transition process due to the lack of readiness and skill of coders to use new classification. Therefore, as a part of the project of ICH, a study on the appropriate approach for computer-assisted coding of the new classification was performed. The aim of this study is finding the suitable and practical model of decision to select the most appropriate code from new classification for each medical procedure or intervention. The model should be implementable in order to develop a computer-assisted coding system.

2. METHODS

Efforts to develop a national classification of health intervention in Iran were started from October 2015 by support of Ministry of Health. Simultaneously, required studies were conducted for choosing the proper method or the decision model of computer aided coding. The first step of the study, we did a review of the existing systems. For this, review measures and comparison criteria were defined in four categories based on type of systems, include general specifications for all of them, specification of fully automated systems, specifications of semi-automated systems and finally for look up systems. According to defined measures, a checklist was prepared. Content of checklist reviewed and verified by two group of experts from field of health information management (four people) and medical informatics (two people). We did search by keywords of “Computer-Assisted Coding”, “CAC software”, “CAC tool”, “medical coding software”, “clinical coding software”. In addition, snowflake search was done from documents, which in some systems were introduced. The search continued until reaching the saturation level. Data collection was done by review of their developers’ websites, catalogs and brochure, web-demo and if possible, observation of trial version of them. Systems were included in the study, which was obtained more than 70 percent of the necessary information about them. We used descriptive analysis for all variables and also Chi square test for some cases.

In the second step, feasibility of each one of CAC types examined with regard to their prerequisites for their implementation and status of infrastructures, medical documentation and coding in Iran. Then, proper model was proposed based on the results of previous steps and the structure of the classification scheme. The model implemented an as interactive system with ASP.NET.

3. RESULTS

3.1. Review of current systems

In the first section, we answer to this question that what is the status of existing coding-assisted systems. It helps to know that what the capabilities and specifications in each of the system types are typical and what prerequisites are essential to their development. Totally 41 CAC systems were identified by searching. Of these, 16 systems were excluded due to more than 70% of the required information about them was not accessible. So, 25 systems were examined. The distribution of them according to general specifications and three types has been depicted in Table 1 (15-42).

There was not a significant relationship between mode of delivery and assistance level (P value ≤ 0.05, Chi-square test). In contrast, a significant relationship was observed between the assistance level and interoperability. So that, all of the fully automated coding systems integrated to Electronic Medical Record (EMR) (P value ≤ 0.05, Chi-square test). In addition, the relationship between method of code assignment and assistance level was examined, but no significant relationship was not observed (P value ≤ 0.05, Chi-square test).
Table 1. General Specifications of studied Computer-Assisted Clinical Coding Applications

<table>
<thead>
<tr>
<th>CAC Types</th>
<th>General Specifications</th>
<th>NLP-Based (N=11)</th>
<th>Semi-Automated (N=4)</th>
<th>Look up Code (N=10)</th>
<th>Total (N=25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delivery Mode</td>
<td>Web-based</td>
<td>(4) 36.4%</td>
<td>(1) 25%</td>
<td>(7) 70%</td>
<td>(12) 48%</td>
</tr>
<tr>
<td></td>
<td>On-premises</td>
<td>(2) 18.2%</td>
<td>(2) 50%</td>
<td>(1) 10%</td>
<td>(5) 20%</td>
</tr>
<tr>
<td></td>
<td>Cloud-based</td>
<td>(5) 45.5%</td>
<td>(1) 25%</td>
<td>(2) 20%</td>
<td>(8) 32%</td>
</tr>
<tr>
<td>Interoperability</td>
<td>Integrated</td>
<td>(11)</td>
<td>(2) 50%</td>
<td>0</td>
<td>(13) 52%</td>
</tr>
<tr>
<td></td>
<td>Outpatient</td>
<td>(4)</td>
<td>(1) 25%</td>
<td>(6) 90%</td>
<td>(10) 40%</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>(13)</td>
<td>(1) 10%</td>
<td>(2) 8%</td>
<td>(15) 60%</td>
</tr>
<tr>
<td>Scope</td>
<td>Hospitals/ Academic Centers</td>
<td>(2) 18.2%</td>
<td>(1) 25%</td>
<td>0</td>
<td>(3) 12%</td>
</tr>
<tr>
<td></td>
<td>Physician office</td>
<td>0</td>
<td>0</td>
<td>(1) 10%</td>
<td>(1) 4%</td>
</tr>
<tr>
<td></td>
<td>Other Health-related Facilities</td>
<td>(1) 9.1%</td>
<td>0</td>
<td>(1) 10%</td>
<td>(2) 8%</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>(8) 72.7%</td>
<td>(3) 75%</td>
<td>(8) 80%</td>
<td>(19) 76%</td>
</tr>
<tr>
<td>Users</td>
<td>Code List</td>
<td>(6) 54.5%</td>
<td>(3) 75%</td>
<td>(10) 100%</td>
<td>(19) 76%</td>
</tr>
<tr>
<td></td>
<td>Single Code</td>
<td>(5) 45.5%</td>
<td>(1) 25%</td>
<td>0</td>
<td>(6) 24%</td>
</tr>
<tr>
<td></td>
<td>North America</td>
<td>(11)</td>
<td>(3) 75%</td>
<td>(8) 80%</td>
<td>(20) 80%</td>
</tr>
<tr>
<td></td>
<td>Europe</td>
<td>0</td>
<td>0</td>
<td>(1) 10%</td>
<td>(1) 4%</td>
</tr>
<tr>
<td></td>
<td>Asia-Pacific</td>
<td>0</td>
<td>(1) 25%</td>
<td>(1) 10%</td>
<td>(2) 8%</td>
</tr>
</tbody>
</table>

32% of systems, we’re able to reimbursement coding addition to the clinical coding. 24% of studied systems have reimbursement coding, coding audit and reporting capabilities. 8% have only reimbursement coding and coding audit, 8% only audits, and reporting, 4% only reporting, 4% only education, 4% only auditing. 8% of systems have all of the above capabilities.

The majority of them (60%), provide the related coding guidelines to users. Seven system (28%) provide medical dictionary as an add on module. Five of the systems (20%) provide other tools such as RVU calculator. 4% have a Clinical Document Improvement (CDI) add on module for medical documents quality auditing or review from aspects of completeness, timeliness and etc. In addition, 4% of studied CAC solutions have both CDI and statistical dashboard (visual statistical reporting tool) as their add on modules. The remaining systems (44%) were not any add on functionality separate from their core modules.

Of the total 11 systems, fully automated or NLP-based, 27.3 percent have used machine-learning techniques, 27.3 percent from a combination of statistical techniques, machine learning and symbolic NLP have used for the processing of medical documentation and coding. The technique used in natural language processing engine was uncertain in 45.5 percent of NLP-based CAC systems. Only 27.3 percent of NLP-based systems allow crosswalk or mapping between used different classification codes and others (73 per cent) had no such feature. More than half of the coding systems based on NLP (55 percent) provide the possibility to edit codes.

Of the total four semi-automated systems examined, 50 percent, guide the coder through a wizard to allocate an appropriate code. Half of semi-automated systems (50%) provide the possibility of code edit to users. 50% of semi-automated coding systems provided crosswalk between related codes of various classifications. Another type of studied systems were look up systems that the distribution of their specific capabilities are presented in the Table 2.

<table>
<thead>
<tr>
<th>Capability</th>
<th>Type</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search</td>
<td>Keyword</td>
<td>(2) 20%</td>
</tr>
<tr>
<td>Code</td>
<td>(1) 10%</td>
<td></td>
</tr>
<tr>
<td>Both</td>
<td>(7) 70%</td>
<td></td>
</tr>
<tr>
<td>Navigation in classification</td>
<td>Drill-down</td>
<td>(6) 60%</td>
</tr>
<tr>
<td>Tree structure</td>
<td>(4) 40%</td>
<td></td>
</tr>
<tr>
<td>Navigation in content</td>
<td>Quick link for exclude codes</td>
<td>(1) 10%</td>
</tr>
<tr>
<td>Quick link for cross references</td>
<td>(2) 20%</td>
<td></td>
</tr>
<tr>
<td>Quick link for cross references and exclude notes</td>
<td>(1) 10%</td>
<td></td>
</tr>
<tr>
<td>Quick link for rubrics and code titles</td>
<td>(4) 40%</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>(2) 20%</td>
<td></td>
</tr>
<tr>
<td>Mapping or crosswalk</td>
<td>Yes</td>
<td>(9) 90%</td>
</tr>
<tr>
<td>No</td>
<td>(1) 10%</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Capabilities of look up CAC systems

3.2. Selection of Feasible CAC Type

In the section, we answer to the question that which types of systems can be developed with regard to the prerequisites and existing infrastructure or conditions. As results of previous step shows, there is a significant relationship between assistance level and interoperability. It is quite certain that, full-automated systems that always created on base of NLP should have access to electronic documents. However, in Iran, electronic medical record hasn’t been implemented literally in hospitals and often is in stage 2 of the EMR adoption model (EMRAM) (43–44). Electronic data are restricted to administrative-financial data, laboratory data and medical images. From clinical data, only some of the main data include diagnoses, surgical procedures, medications, monitored vital signs and discharge data as structured entered into the hospital information system. The narrative data are very limited in electronic format and confined to radiology or pathology reports.

In some of hospitals, paper medical records scanned and stored electronically. However, they must be processable with optical character recognition (OCR) technique. In addition, the current language of medical documentation is Persian, so that physicians and nurses usually write their notes, except radiology and pathology reports, in the Persian language but also use of medical terminologies and abbreviations in English in their notes. Although there are a few physicians who write medical notes in English, however, there are no rules about documentation language. In such situations, proper CAC can’t be built up in the form of a fully-automated or NLP-based. According to the results of the previous step, it seems that look up systems due to their navigation and search capabilities more used to verify the selected code. So it would not be appropriate in situations where coders are faced with a new classification and even, they do not know how to choose a code. Therefore, a semi-automated CAC is more suitable.

3.3. Proposed three-phased decision model

In the section, we respond to this question that what is appropriate model to develop a semi-automatic system to help code selection and to facilitate coding with the new classification. We proposed a model based on the classification scheme of IRCHI, which is in accordance with the ICHI classification scheme in structure. In this classification scheme, each intervention is specified and classified based on three axis include “Action” (that is, work that’s done), “Target” (the entity on which the Action is performed) and “Means” (approach,
A Three-Phase Decision Model of Computer-Aided Coding for the Iranian Classification of Health Interventions (IRCHI)

The model designed based on hierarchical relationships between entities in the classification scheme and also the logic of decision making to specify the characters of code step by step through a web-based interactive user interface for CAC. It is composed of three phases: (i) select Target, (ii) select Action and (iii) select Means. For example, Figure 1 shows the logic of decision making to determine the code of “open pupillotomy”.

In general, to implementation of the model, we identified valid relations that generate valid codes and defined the rules specifying code characters based on the question and user response and selection of possible entities from related options in each phase. The number of questions in the first, second and third phase of decision model is four, three and two respectively. Figures 2-4 show function of designed CAC solution for the example of “open pupillotomy” based on presented model.

4. DISCUSSION

As results of review on existing systems shown, fully automated CAC systems, which are on base of NLP, are popular solutions. But, despite their popularity, they need to some infrastructures such as integration with electronic clinical documents, using the standard terminologies and unique language for documentation. Our results show a significant relationship between the assistance level and interoperability with HIS or EMRs. As Jusinski mentioned that NLP based CAC systems are dependent on fully electronic medical documents (45). As regards, no hospital in Iran, do not have electronic medical records and studies have been shown the majority of hospitals are in stage 2 of EMRAM (43-44), implementation of fully automated CAC systems is impossible. In addition, several researchers emphasized on the necessity of using the unique language and standard terminology in medical documentation for implementing NLP based systems (46-48). However, in Iran, physician and other care specialists use a combination of both languages of English and Persian for clinical documentation (49). This problem the part of speech tagging makes difficult to develop the NLP-based CAC systems.

One of the interesting finding of our review on existing CAC systems was that mostly NLP-based systems had been developed in North American countries and other countries mostly turn to develop non-NLP based or semi-automated CAC systems. Our results can be compatible with the results of other studies which believe that the development of NLP-based systems is difficult to agglutinative languages or synthetic languages that use morphological agglutinative such as German, Dutch, Arabic and Persian (46, 50-52). According to Xuehelaiti, agglutinative languages widely used in Korea, Japan, Turkey and other countries in Middle East and Asia (53). These languages have complex morphological phenomena; consequently, morphological processing is important and difficult in NLP while the NLP systems that were created focused on word processing (50-51).

Although, the results have been shown that semi-automated systems are less popular than the other types of CAC
systems, but, they are more practical for countries with similar status of Iran from aspect of infrastructure and language. Fortunately, hierarchical structure in the classification scheme of the new classification, allowed developing the suitable model of decision-making for selecting the related code in step-by-step way. The study of Ning demonstrated how the hierarchical structure of a classification could help to improve effectiveness of CAC system (54). Of course, the effectiveness of this model on choosing the most correct code was not evaluated which is a limitation of our study. Therefore, one of the important next steps that must be done in the continuation of this study is to evaluate the system. Evaluate the effectiveness of the system on the correct coding with new classification and usability of it for users. In addition, to determine the correct Action Detail or Action Title of an intervention may need to set more rules and add further questions to the second phase of the model. It is possible through if-then rules with regard to the provided definitions and inclusion or exclusion notes for each Action Titles in the classification.

5. Conclusion

The proposed model that designed according to required infrastructures for each type of CAC solution and the current stratus of clinical documentation and coding in Iran and also, the structure of new classification scheme. Our results show it was practical. However, the model needs to be evaluated in the next stage of the research.

A Three-Phase Decision Model of Computer-Aided Coding for the Iranian Classification of Health Interventions (IRCHI)

Acknowledgements: This study is a part of common research between Ministry of Health and Medical Education of Iran, (Deputy of Treatment, office of statistics and treatment information management) and Tehran University of Medical sciences as PhD thesis. Financial support provided by Ministry of Health. We thank from Ms. Akram Vahedi and other persons who provided support and expertise that greatly assisted the research.

Conflict of Interest: none declared.

REFERENCES

A Three-Phase Decision Model of Computer-Aided Coding for the Iranian Classification of Health Interventions (IRCHI)


