

# Standard Chief Complaint Set Created from Discharge Summary, Applicable to Electronic Medical Record: Short-term Experience in Seoul National University Bundang Hospital

Ho Jun Chin<sup>1</sup>, Choong Ho Shin<sup>2</sup>, Jun dong Park<sup>2</sup>, Woo Young Jeong<sup>3</sup>, Kyoo Seob Ha<sup>4</sup>, Nam Soo Byeon<sup>5</sup>,  
Suhnggwon Kim<sup>1</sup>

<sup>1</sup>Department of Internal Medicine, <sup>2</sup>Department of Pediatrics,  
Seoul National University College of Medicine, Seoul, Korea,  
<sup>3</sup>Department of Internal Medicine, <sup>4</sup>Department of Psychiatry,  
Seoul National University Bundang Hospital, Seong-nam, Korea,  
<sup>5</sup>ezCaretTech, Seoul, Korea

## ABSTRACT

**Introduction :** Seoul National University Hospital implemented the first version of EMR to new-branch hospital in Seong-nam, Korea. To develop a clinically useful and standardized set of chief complaints, we extracted chief complaints from discharge summaries in Seoul National University Hospital and applied it to EMR.

**Method :** First of all, we parsed chief complaints as main concept and attribute. We normalized main concepts and selected term set. And then, mapping the term to SNOMED CT through one by one matching process was done. We analyzed the utilization status of coded chief complaint set for 50 days.

**Result :** Among 94,913 unique strings extracted from 220,200 discharge summaries, we initially selected 6,317 terms as initial standard set of chief complainst. EMR users evaluated the set before operating EMR system and added 431 terms to the final set. The majority of terms in the final set were exactly mapped to SNOMED CT conceptually. The usage rate of coded set was 47.3% out of recorded 21,678 chief complaints in doctor's notes. But, only 9.1% of 11,419 chief complaints, inputted in text, were new strings to our set. Finally, they were normalized to supplement 437 new terms to standard set.

**Conclusion :** We developed standard chief complaint set, mapped to SNOMED CT. Users of EMR tended to chart chief complaints in text, which were already included in coded set. For recording chief complaints as useful coded data, more attention and discussion about users' traditional practice should be paid.

**Key words :** standard chief complaint, EMR, SNOMED CT

## INTRODUCTION

Electronic medical record (EMR) technology is spreading into world wide healthcare environment, including Korean medical society. Many papers report potential benefits of EMR but efforts, such as structured data input with controlled medical vocabulary, standardization, and system integration are required to harvest benefits[1]. We need domain specific and user-friendly medical terminologies with which user can effectively capture clinical events and

manage clinical data in certain institutional environment. But, there are no noble standard terminologies capturing details of clinical problems, diagnoses, or procedures with sufficiency to satisfy users' need until now. Some terminologies are outstanding but they still have unfamiliar aspects to users, especially whose mother language is not English.

Seoul National University Hospital (SNUH) adopted computer system in 1979 for only reimbursement and management. In mid 1990's, large tertiary hospitals in Korea turned their interest in hospital computerization to clinical fields and began to develop many hospital information systems, such as physician order entry system (POE) and PACS. SNUH has implemented POE system and PACS five years ago, too. Now, the hospital has opened new branch hospital, Bundang SNUH, which has 800 beds for inpatient care and 6 centers and 31 clinical departments for outpatient care. The new version of EMR is implemented to this hospital as well as PACS and POE on May this year. The hospital achieved paperless, radiology-filmless, prescription-slipless, and medical-chartless environment in hardware system.

SNUH has indexed patients' diagnoses and procedures with local code system, mapped to ICD-10 or EDI code for reimbursement. For new hospital, we are trying to standardize patients' chief complaints and nomenclature of drugs along with extension of diagnosis and procedure term set. The chief complaint is directly collected by physicians and represents primary information that directs medical care. Also, it is sort of fundamental source of information on which computerized decision support systems, such as guidelines, protocols, or critical pathways, are based. We did not adopt general-purpose standard terminologies, such as UMLS or SNOMED CT as it is not only because they include too huge and complicated contents to use on the point of care but also they certainly don't have Korean medical terms. On the contrary, we choose the way to normalize our data in SNUH and map them to standard terminology. We think it is more domain-specific and user-friendly method to develop controlled chief complaint set relevant to our institution.

In this study, we tried to develop a user-friendly and standard set of chief complaints, to map it to general standard terminology, and to evaluate user's habit in recording chief complaints.

## METHODS

### Selection of standard set

SNUH constructed discharge summary data base of all clinical department except psychiatry in 1999 and recorded chief complaint at admission in text. The discharge summaries were collected for about 7 years, counted 235,426. From data base, twenty seven medical experts selected the chief complaint set. First of all, researchers parsed chief complaints as main concept, qualifier, and modifier. And then, we normalized the main concept following process. The normalization orders the words in strings alphabetically, eliminates article words, prepositional words and postpositional words, and converts the words to lower case if they are English and to standard canonical terms. We analyzed their concepts and compared their strings. If some words had same concept and different strings, we chose one typical string and consider the others as synonyms. Among 94,913 unique strings, we selected 6,317 terms as initial standard set of chief complaint. The users of EMR were asked to evaluate usability of the set and add some words if they needed. Supplemented strings were normalized, analyzed their concepts and compared to initial standard set. New concepts were included in the final standard set, being composed of 6,748 terms.

### Mapping the set to SNOMED CT

The chief complaint domain includes symptoms, signs, diagnoses, procedures, or anatomical structures in concepts. When we decide which standard vocabularies are suitable to map chief complaint set to, we give the priority to comprehensiveness of vocabularies. SNOMED has more completeness and richer taxonomy, especially in finding, diagnosis, and intervention category[2,3,4]. SNOMED CT, new version of SNOMED released on July 2002 which was merged with READ code and got more abundant concepts. Four physicians mapped a normalized term to single concept identifier of SNOMED CT and one physician of them finally confirmed consistency and accuracy of mapping process. If one term in normalized set matched to multiple fullyspecifiednames in snomedconcepts table of SNOMED CT, we selected the most relevant concept to use in chief complaint domain. Thus, we considered only one matching SNOMED concept per a term from normalized set. We analyzed mapping status as identical mapping which means a SNOMED concept and a normalized term are identical correlates in concept, parent mapping which means normalized term is more specific than SNOMED concept, child mapping which means normalized term has more broader concept than SNOMED, overlap mapping which means overlap exists between correlates, and unmappable. And we classified normalized set as hierarchical groups of

SNOMED CT.

### Analysis of user's practice

Bundang SNUH opened at May 10<sup>th</sup> this year. We evaluated user's attitude to coded data entry at 50 days after implementing EMR. We selected doctors' notes in outpatient and inpatient setting and searched chief complaint fields. We normalized again text data and compared them to standard chief complaints.

### Chief complaint input interface in EMR

User can record chief complaints in code or text. We provide search browser to find coded data and present another fields in which user can input attribute, such as qualifier or modifier to explain chief complaints in detail (figure 1). Users can also compose the attribute set to be selected automatically with relationship to chief complaints. Among 31 clinical departments, fourteen have attribute sets. Four departments made their attribute sets specific to chief complaints and the others made common attributes.

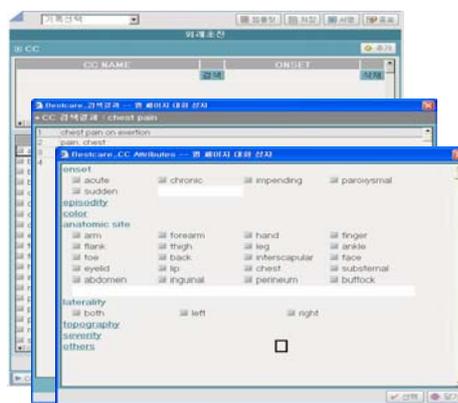


Figure 1. Input interface for chief complaint

## RESULTS

### Standard chief complaint set

Discharge summaries for 7 years in SNUH were 235,426 and chief complaint fields were evident in 220,200 discharge summaries. After sorting 220,200 chief complaint fields alphabetically and parsing main concepts, we could find out 94,913 unique strings. We normalized them alphabetically and conceptually and selected 6,317 main concepts. Users evaluated the initial set and requested to add 3,480 strings. We analyzed those strings and figured out only 431 terms as new items to initial set. We finalized 6,748 terms as standard chief complaint set, in which psychiatric chief complaints were not included.

### Mapping the set to SNOMED CT

The standard chief complaint set was mapped to 5,361 unique concepts of SNOMED CT. Among 6,748 terms, 91.4 percents were mapped to identical concepts. Parent mapping was 7.0 percents, child mapping 0.2 percents, overlap mapping 0.9 percents. We could not find related concepts in thirty one(0.5%) chief complaints. They were

given new concept identifier and parent relationships with SNOMED CT concepts by researchers. We categorized chief complaints grouped by fullyspecifiedname in SNOMED concepts file. The finding category enclosed 29.8 percents of terms, disorder 47.8 percents, procedure 16.2 percents, and others, such as morphologic abnormality, tumor staging, life style, context dependent category, qualifier value, and function, 6.2 percents.

### User's habit of practice

For 50 days after implementation of EMR, users wrote 19,105 doctor's notes at first confrontation of 10,628 patients. Doctors recorded one to six chief complaints in one note (mean number of chief complaint in one note is 1.13). Forty-seven percents of recorded 21,678 chief complaints were inputted in coded terms. The chief complaints written in text had been almost included in standard set. Only 1,037 strings (9.1%) among 11,419 text strings were new to the set. After normalization alphabetically and conceptually, we figured out 437 terms as new chief complaints. They were only 6.4 percents of standard terms.

## DISCUSSION

In this study, we created a standard coded chief complaint set for EMR. It was based on actual data of our institutional discharge summaries. We did not aim to input chief complaints in code completely. It is important to give the chance to users to record in free text because changing documentation practice from free text to coded data entry requires time[5]. Free-text entry also provides opportunities to detect users' requirement in chief complaint domain and improve the set. At the same time, free-text entry compensates to express delicate meaning by local language in addition to code. Korean language has abundant adjective expression. Chief complaints that could not be categorized and coded are inevitably remained, especially if the code is based on English. If we insist to remove free-text data entry in EMR, it will result in inaccurate data input or user's indifference to coded data entry.

In selection of standard set from clinical data, granularity is one of common problems[5]. We did not compose the set in same granularity because we pursued to improve real time documentation and to satisfy actual user's need. Each departmental expert selected the most relevant terms in statistically analyzed data pool to chart main concepts of chief complaints rapidly and conveniently. If the set is in high granularity, ambiguous expression by patient could not be coded at all because patient's expression is not always specific. The conceptual granularity is also influenced by certain clinical settings. Time limitation is one of main obstacles to chart in detail. In outpatient clinic or emergency room, physician has usually less than 10 minutes to provide medical services to patients at first confrontation in Korea. Encounter history is another factor. The granularity of chief complaints would be different in recurrent scheduled admission from in the first admission. Departmental difference was noticed in this study. Doctors in

ophthalmology, psychiatry, and dentistry used Korean language more frequently and more ambiguous terms than any other departments.

Korean doctors use English in describing chief complaints in general. In discharge summary database, Korean was less than 5 percents except in ophthalmology, dentistry, and pediatric orthopedics. Chief complaints described in words or word phrase were simple to normalize. The problem was the adjective expression in Korean without clear concepts. Users tended to record such ambiguous expression as sentences in Korean instead of words in English or Korean. Users thought that way was appropriate to describe patients' condition. We gave codes to main concepts of such sentences but users did not accept those codes. So, we concluded free-text entry would be more relevant to express delicate and ambiguous meanings in Korean sentence.

Almost concepts in standard set were completely matched to SNOMED CT. It was because we selected only main concepts to describe chief complaints rapidly and conveniently. We composed another interface to record attributes which were dissected from normalized terms. SNOMED CT is excellent to present concepts in finding, diagnosis, and procedure domain[2,3,4] which were main parts in our set. We did not permit compositional mapping. If we did, matching rate would be higher. For example, skin lesion of scalp was given a new identifier but considering post-coordinate mapping, it would be expressed as following:

Skin lesion of scalp :

(is a: *conceptid=116680003*)

skin lesion: *conceptid=95324001*

(*finding site : conceptid=363698007*)

entire scalp : *conceptid=181469002*

Our set covered over ninety percents of concepts used for 50 days after implementation of EMR. But, users charted with coded data less than fifty percents. Doctors thought differently each other about significance of chief complaint as data. While discussing the principle of data input, medical doctors considered chief complaint to be recorded in code but surgeons did not. Surgeons were more interested in procedure and diagnosis domain. In our survey, frequency of free-text entry in surgical parts was 1.6 times higher than that of medical parts. To our surprise, over ninety-five percents of chief complaints recorded in neurology and dermatology department were coded data. On the contrary, seventy-two percents of chief complaints in six major surgical departments were recorded in free-text. This phenomenon also related to the content of standard chief complaints. The seventy-six percents of chief complaints which should be added to the set were from surgeon's notes in spite of the number of patients in both medical and surgical department being similar. The leader of standardization project was medical doctor and enthusiastic participants were also medical doctors. So, there would be unbalance to choose terms in surgical fields. In emergency room, doctors of various departments recorded chief complaints in free text but most of their terms were simple English words found in standard set. That was because

doctors who confronted patients at first were residents sent from mother hospital for short time except three house staffs. They did not have enough time to be skilled at EMR and coded data entry. Education of transient users is one of major problems to encourage proper usage of EMR till now. Another problem is adaptation time to EMR environment. Only a few staffs participated in EMR project from beginning. Many were exposed to EMR less than four months. All of them had no experience to use EMR and had different experience in hospital information systems. In Aronsky's report[5], users need more than 2 years to fully adopt coded data entry. Finally, users did not benefit by coded data entry yet. It has been just potential till now. Data storage and data analysis are just started. If we could show the benefit to them, they would be more encouraged to chart in code.

This study has several limitations. As previously mentioned, standard set was for EMR and composed by potential users. We did not try to standardize all chief complaints in discharge summaries of SNUH. If we want to set chief complaints applicable to any institution, more effort is needed. But our approaches must be also effective to get that goal. In mapping terms to SNOMED CT concepts, one to one matching was essential. If we considered post-coordinate mapping, matching result would be different from this report. And we did not evaluate appropriateness of SNOMED terms themselves. That was out of our scope. We did not directly evaluate all users' attitudes to coded data entry. We just estimate them by analyzing data of chief complaints, departments, users' names, and identifiers of doctor's notes recorded for 50 days. We did not include psychiatric terms in chief complaint set. We did not assess the effectiveness and convenience of user interface to record chief complaints. We will study those themes near future.

In conclusion, we successfully installed standard chief complaint set to EMR. It covered over ninety percents of recorded concepts in chief complaint fields. But, we need more attention and discussion about user's attitude to coded data entry for increasing usage rate of standard chief complaint set. Moreover, we must revise and reevaluate the set continuously to satisfy user's requirement.

#### ACKNOWLEDGEMENT

We are thankful to all participants in standardization project and related persons to EMR development. Especially, we appreciate ezCareTech's support and funding with all my heart.

#### REFERENCES

- [1] Powsner SM, Wyatt JC, Wright P. Opportunities for and challenges of computerization, *Lancet* 1998;352;1671-22
- [2] Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR for the Computer-Based Patient Record Institute's Work

Group on Codes & Structures. The Content Coverage of Clinical Classifications, *JAMIA* 1996;3;224-233

[3] Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG, Warren J for CPRI Work Group on Codes and Structures. Phase II Evaluation of Clinical Coding Schemes: Completeness, Taxonomy, Mapping, Definitions, and Clarity, *JAMIA* 1997;4;238-251

[4] Humphreys BL, McCray AT, Cheh ML. Evaluating the Coverage of Controlled Health Data Terminologies: Report on the Results of the NLM/AHCPR Large Scale Vocabulary Test, *JAMIA* 1997;4;484-500

[5] Aronsky D, Kendall D, Merkley K, James BC, Haug PJ. A Comprehensive Set of Coded Chief Complaints for the Emergency Department, *Acad Emerg Med* 2001;8;980-989

Address for correspondence

Suhnggwon Kim, MD, PhD

address : Division of Nephrology,  
Department of Internal Medicine,  
Seoul National University Hospital,  
Yeon-gun dong 28, Jong-ro Gu,  
Seoul, Korea, 110-744

e-mail : skimim@plaza.snu.ac.kr