

**MEETING OF WHO COLLABORATING CENTRES
FOR THE FAMILY OF INTERNATIONAL CLASSIFICATIONS**

Tunis, Tunisia
29 Oct. - 4 Nov. 2006

Process analysis on health care episodes by ICPC-2

Shinsuke Fujita¹⁾²⁾, Takahiro Suzuki³⁾, Katsuhiko Takabayashi³⁾.

- 1)WONCA International Classification Committee
- 2)Department of Welfare and Medical Intelligence,
Chiba University Hospital,
- 3)Department of Medical Informatics,
Chiba University Hospital.

Japan

Abstract

[Back ground] IT brought electronic medical records and IT will also bring electronic health records in the near future, which means we will be able to analyze a large amount of health data. In another words, we have to develop the way to analyze EHR from the stand point of public health and as for personal health improvement we have to develop the way to find critical health episodes from the life-span health record. [Purpose] In this study we want to develop extracting method of health care episodes from EHR. This technique will help us to analyze EHR data for public health. We also want to show the example analysis method for public health. [Materials and Methods] We have full set of medical records for more than three years in Chiba University Hospital. We extracted all records of colon cancer and mapped health care episodes, such as symptoms, examinations, prescriptions, interventions and diagnoses, onto ICPC-2. In symptom mapping, i.e., chapter 1 of ICPC-2, we used Japanese language morphological analysis technique. We could map all other care episodes onto chapter 2 to 7 without difficulty. After these preparations we did morphological analysis again to make sub classification of these patients according to appearance similarity of ICPC-2 codes, which we hoped to distinguish the purpose of admission. Then we compared the sequence of care episodes. [Results] We could successfully distinguish the purpose of admission by morphological analysis. Sequence analysis suggested too much blood tests or x-ray examinations in some patients. [Conclusion] ICPC-2 was developed to use in primary-care setting and it was also aimed to international comparison. For these purpose, it was designed to be an adequately rude classification. Thanks to this adequate granularity, ICPC-2 is very convenient for health care episodes analysis even in hospital or in public health. This result suggests that we can use ICPC-2 to improve national health status.

This document is not issued to the general public, and all rights are reserved by the World Health Organization (WHO). The document may not be reviewed, abstracted, quoted, reproduced or translated, in part or in whole, without the prior written permission of WHO. No part of this document may be stored in a retrieval system or transmitted in any form or by any means - electronic, mechanical or other - without the prior written permission of WHO.

The views expressed in documents by named authors are solely the responsibility of those authors.

Content

Abstract 1

Background 3

Study design 3

 Patient record extraction 3

 Patients list 3

 Symptom table 4

 Test table 4

 Prescription table 4

 Intervention table 4

 Pilling up processes in each admission 4

 Statistical analysis of words appearance 4

 Comparison of ICPC-2 code lists and admission summary 5

Results 5

 Extraction of symptom 5

 Participation of ICPC-2 code list for process analysis 5

 Participation of ICPC-2 code list for summary making 5

Discussion 6

Background

In the last ten years, IT brought us medical record from paper form to electronic form (EMR). We have used full set of EMR in these three years in our hospital. We made questionnaire in our institute whether we should go back to paper based medical record. No one hoped to go back and we regard that EMR is useful in daily practice. In that questionnaire, however, the difficulty in reference of previous describes was raised. In case of paper medical record, the thickness tells us the time course of patient's history and we need some improvements in EMR. They also pointed out the necessity of summarization. We make summaries when we discharge our patients and we realized that we have to summarize medical record also in outpatients. Moreover, some attempts to expand EMR to national wide level or life long time, which is cold electronic health record, "EHR". From technological aspect, we are now able to building up EHR. In an emergency room, easy accessibility to whole of personal medical record will bring important information, but is it able for clinicians to read all through the patient's life long medical record? The more information we can get, we need more good method of picking up important information, which means we need good summarization method. In this paper, we aimed to develop summarization method and to analyze medical process with help of ICPC-2.

Study design

We tried to summarize medical record, which is written in daily practice in natural language.

Patient record extraction

In this study, we set colonic cancer as target disease.

Patients list

We made patients list from EMR, that is, all patients who were diagnosed ICD-10:C18.* between January 2004 and December 2006. This patients list table consists with patient's ID, patient's name, ICD-10 code, the date of diagnosis, the date of admission, and the date of discharge. We did not included secondary malignant neoplasm of large intestine (C78.5).

If one patient has more than two diagnoses, in the same admission, we packed them in one record. It means every double-cancer-bearing patient in the same admission has only one record in our patients list.

If one patient has more than two admission and who has C18.* in each admission, we kept each record.

Symptom table

We use SOAP system in describing medical record and we extracted all records from symptom field (S-field) according to the patients list. S-field is written in free text. Symptom table consists with patient's ID, the date of admission, the date of discharge, the date of describing, and S-field.

We applied Japanese language morphological analysis technique with "Chasen" on S-field. The analyzed word by word separated sentences were then applied dependency structure analysis with "CaboCha". For example, "He dose not have abdominal pain, nausea, nor constipation" is "'no-abdominal-pain', 'no-nausea', 'no-constipation'". These processes prevent coding to D06, D09, and D12. The final step we applied ICPC-2 coding with Japanese ICPC-2 dictionary.

Test table

According to the history of tests' order during targeted admission, we created test table. We made ICPC-2 coding by the kind of test order and we did not concerned about the organ related chapter of ICPC-2. Test table consists with patient's ID, the date of admission, the date of discharge, the date of examination, ICPC-2 (*32 - *43), and the name of test.

Prescription table

We coded all executed prescriptions and injections to *50. Prescription table consists with patient's ID, the date of admission, the date of discharge, the date of drug intake, ICPC-2 (*50), and the name of drug.

Intervention table

We coded all executed major surgery to ICPC-2 (*52), though this coded is intended for minor surgery. We coded *51, *53, and *54 as in original meaning. Intervention table consists with patient's ID, the date of admission, the date of discharge, the date of intervention, ICPC-2, and the name of intervention.

Pilling up processes in each admission

We pilled up symptom, test, prescription, and intervention tables into one process table. Process table consists with patient's ID, the date of admission, the date of discharge, the date of execution, ICPC-2, and item name. With this database, we performed following studies.

Statistical analysis of words appearance

From pilled up database, each admission data were extracted. In these extractions ICPC-2 codes were lined up in time sequence and we applied text-mining techniques.

Comparison of ICPC-2 code lists and admission summary

The same ICPC-2 code extraction data were used for this study.

Results

Targeted patients (ICD10:C18*) were 318 patients. The purpose of admission were surgical treatment, chemotherapy, and to make definite diagnosis. Free-text written symptoms were successfully coded to ICPC-2. We could deal both positive symptom lists and negative symptom lists. Dependency structure analysis

Extraction of symptom

Extracted ICPC-2 codes from symptom field were visually check with original text. Japanese language morphological analysis perfectly extracted key words that fit to ICPC-2 dictionary. Due to listed symptoms in one sentence, and which sometimes in negative, we employed dependency structure analysis. Thanks to this process, we could drop negative symptoms from ICPC-2 coding and achieved 83% accuracy.

Participation of ICPC-2 code list for process analysis

Pilled up ICPC-2 code list shows symptom oriented i.e., episode based clinical course. From this list we can correctly guess the purpose of admission (88% accuracy). Excessive blood test repetition before surgery suggests some kind of complication. This estimation was 78% accuracy and all the rest were evaluated as not necessary tests.

Participation of ICPC-2 code list for summary making

From the piled up ICPC-2 code list, all repetitions of ICPC-2 codes were excluded. Because ICPC-2 cannot deal the difference of imaging tests such as CT, MRI, and Ultrasonography, the name of modality was dealt together with ICPC-2 code. Surgical treatment was dealt in the same manner. The extraction was compared with admission summary.

According to the purpose of admission, we got different results. For surgical treatment and for definite diagnosis, we got fairly good information. Histological diagnosis and staging of malignancy, and the plan after discharge are the information that we could not get from this study and that we can get easily from EHR. For chemotherapy, we need the name of protocol and the stage of completion, the present status of health (clinical findings and tests), and efficacy of the treatment but we could not get these kinds of information from our study.

Discussion

In information technology, the most important thing is to utilize data. For which purpose we have to classify data and the granularity of classification is very important to analyze data. However little study have made against the granularity. In this study we selected ICPC-2 to analyze data. ICPC-2 was designed to analyze clinical data and process. Its target is for primary care and not for hospital care. To get rid of this difference we used name of imaging modality and name of surgical procedure with ICPC-2 code. This method is easy way to expand rude classification. WONCA international classification committee is now planning to collaborate ICPC-2 with ICF in the same manner.

We also tried automatic coding from narrative description. We employed dependency structure analysis for this purpose. We Japanese and other 2-byte character using country have to use front-end-processor to determine the character we want to use. In the previous report we advocated to include coding on the dictionary of front-end processor. If we use only one classification in any occasion, this method is very useful, but as for data analysis we may have to use several classification according to our purpose. Automatic coding system from natural language will be helpful in that situation and we showed the usefulness of employing dependency structure analysis.