

Evaluation of the CCAM Hierarchy and Semi Structured Code for Retrieving Relevant Procedures in a Hospital Case Mix Database

Cédric Bousquet^{1,2}, PharmD, PhD, Béatrice Trombert^{1,2}, MD, PhD, Julien Souvignet¹, MS, Eric Sadou¹ and Jean-Marie Rodrigues^{1,2,3}, MD, PhD

¹ Department of Public Health and Medical Informatics, CHU Université Jean-Monnet, Saint Etienne, France; ² INSERM UMR_S 872, Eq 20, Paris, France; ³ WHO FIC Collaborative Centre for International Classifications in French Language, France

cedric.bousquet@chu-st-etienne.fr

Abstract

In France, clinical procedures are coded with the French procedures classification (Classification Commune des Actes Médicaux, CCAM) and recorded in every hospital. CCAM uses hierarchical semi-structured codes which describe procedures (topography, action, access mode and/or technique). This amount of information could be analyzed and used for clinical and medico-economic evaluation. But relevant and practical data searches are difficult. In this paper we present a use case about searching for endoscopic activities in a case mix database to evaluate the relevance of the hierarchical organization and semi structured codes of CCAM in order to retrieve data already coded using this controlled vocabulary. Precision was 0.79 and recall 0.84 in the hierarchical search whereas precision was 0.94 and recall was 0.81 using part of the code related to access mode and/or technique. We discuss a revision of the CCAM by the use of an existing modeling (from the GALEN project) and better knowledge representation for each procedure.

Introduction

France uses its own national coding system for surgical and interventional procedures. Named following a French acronym: CCAM (*Classification Commune des Actes Médicaux*) which has been developed since 1996 [1]. It is mainly used in hospitals since 2004 for DRG based funding from case mix databases and fee for service payment for private physicians outside hospitals. The terminology labels have been aligned to be conformant with the European standard on the Categorical Structure for Surgical procedures EN-1828 [2] and populated by the Galen Ontology driven tools [3].

This classification uses a semi-structured code with four letters and three numbers which allows a basic understanding of the multi axial structure by reading the code. It is currently used to search through codes and analyze the encoded data in

medico economic evaluation, for example to estimate the return on investment after buying a new expensive equipment or opening new hospital beds. Practical experience shows that using CCAM in such studies is difficult and it is hard to find all the relevant procedures within all these medico economic studies.

Several innovative methods have been proposed for assisting in coding information in the medical record especially for diagnoses using the International Classification of Diseases [4-6]. However, reporting on the searches for information once it is coded using a controlled terminology and stored in a database has been limited.

Our objective was to evaluate the hierarchical organization and semi structured codes of CCAM for retrieving data already coded with this controlled vocabulary. This work was performed within the French INTERSTIS project (*Interopérabilité Sémantique des Terminologies dans les systèmes d'Information de Santé Français* - Semantic Interoperability of Terminologies in the French Health Information Systems).

In this article, we take as example the medico-economic evaluation of the renewal of endoscopy devices in a hospital. The purchase of endoscopes is conditioned by a study on the number of endoscopic procedures performed in the hospital. In this scenario, the endoscopic department needs to evaluate endoscopic activity in the case mix database. In theory, it should be possible to search all CCAM codes corresponding to actions taken under endoscopy in the nomenclature. Practically the search for concepts within CCAM is incomplete and we consider new ways to identify additional codes. We implemented three search strategies for CCAM procedures as described in the guide from *Groupement de modernisation du système d'information hospitalier (GMSIH)* [7]. These search-modes are by: keyword, hierarchy, partial codes.

Materials

CCAM: An electronic version (CCAM v17) was downloaded from the French national Health insurance website [8]. We worked on Chapter 7 (Digestive procedures) and the number of CCAM codes was 1,155 in this study. New updates are made every year.

CCAM codes have a defined 7-characters structure. The first letter designates the system, tract or structure which the action concerns. The second letter provides additional detail such as a function or an organ. Table 1 is an excerpt from the English translation for anatomical site. The CCAM descriptors are described in the *Guide de lecture et de codage* (Reading and Coding guide) [9]. The letter *H* is related to *Digestive System*. Descriptors describe both anatomical site (e.g. *HC: Salivary glands*) and functions (e.g. *HQ Swallowing*).

DIGESTIVE SYSTEM	H
Lips, tongue, oral cavity as a whole	HA
Teeth, periodontium, gums	HB
Salivary glands	HC
...	H...
Swallowing	HQ
...	H...

Table 1: Excerpt of CCAM coding table for anatomical site and function

The third letter indicates the action. CCAM Action terms have been defined by grouping them according to action type; each is identified by an action verb, and then coded. The same code is often allocated to several action verbs, grouped together because of their technical similarity. Table 2 is an example of actions from the *Guide* [9] that were translated in English. Several actions are grouped under the general category of *Observation Actions*. This general category is associated to several verbs e.g. *Examine* or *Measure*. The first column defines a verb e.g. *Examine*. Several descriptors are associated to each verb and English translations of the descriptors are proposed in the second column (e.g. *Analysis, Exploration, Monitoring*, etc.). The third column shows the letter associated with the verb. The fourth letter indicates the access mode. Table 3 is an example of access mode from the *Guide* [9].

EXAMINE		
observe the body or one of its elements, directly or using instruments, to study or monitor its functioning, without producing a lasting recording	ANALYSIS DETECTION STUDY EXAMINATION EXPLORATION MONITORING SEARCH SURVEILLANCE TRANSILLUMINATION -SCOPY	Q

Table 2: Excerpt of CCAM coding table for actions

CCAM labels are divided into 18 chapters that are first levels in the hierarchical classification. These chapters are identified with 2 numeric characters and defined by the anatomophysiological system. Each chapter is subdivided into sub-chapters, coded with 2 numeric characters. Each sub-chapter is subdivided into paragraphs and some of them are segmented into subparagraphs. The final level is the precise label procedure.

TRANSORIFICE ENDOSCOPIC ACCESS		
access to the lumen of a cavity or anatomical duct via a natural or artificial external orifice (cutaneous ostomy), with the insertion of an optical instrument	by (...) endoscopy by (...) fibroscopy) by (...) laryngoscopy by jejunoscopy by ileoscopy by (...) coloscopy by rectosigmoidoscopy by anterograde ureteroscopy by colposcopy ...	E

Table 3: Excerpt of CCAM coding table for access modes.

Diagnostic procedures are ordered by technique and then by topography. Therapeutic procedures are ordered by topography then by action. Table 4 shows an example of a CCAM procedure with corresponding hierarchy and code.

7	Digestive System
07.01	Digestive system Diagnostic Procedures
07.01.09	Digestive system endoscopy
07.01.09.01	Digestive tract endoscopy
HEQE002	Oeso-gastro-duodenal Endoscopy
H	Digestive System
E	Oesophagus
Q	Examine
E	Transorifice Endoscopic Access

Table 4: Example of code reading.

Methods

We excluded procedures by the means of laparoscopes. Endoscopic procedures are performed in the endoscopic department by using natural body openings whereas a laparotomy is a surgical procedure in the operating room involving an incision through the abdominal wall to gain access into the abdominal cavity. We reviewed the 1,155 digestive procedures and manually selected 121 endoscopic procedures using an expert approach (35 diagnostic procedures and 86 therapeutic procedures). This set of 121 endoscopic procedures was our gold standard.

The GMSIH document recommends three search strategies [7] that we implemented using SQL queries in the CCAM electronic edition. These three complementary CCAM code-search strategies are:

- By keyword
- By procedure position in the hierarchy
- By semi-structured codes

Full-text search: a keyword search can be done on the procedure label. Multiple criteria for words or partial keywords can be used with operators (and, or, not). We searched for CCAM labels containing the following strings: (1) “endoscopy” then (2) “scopy”. Indeed a specialized term is often chosen instead of endoscopy such as gastroduodenoscopy and can be retrieved using the –scopy suffix.

Hierarchical tree search: We reviewed manually the hierarchy of Chapter 7 and selected all branches of the hierarchy that contain at least 50% of codes relative to endoscopic procedures. Then we identified for each branch which procedures are performed by means of an endoscope and which procedures use other means. From this list of branches we selected a subgroup that contains the following strings: “endoscopy” or “endoscopic”. We also performed a search on the hierarchical levels that subsume the branches selected at the previous step.

Code search: Code search can be performed with a partial or complete code, (e.g. a wildcard (% , * , _ , ?) can be used). We selected codes starting by H as

a first letter without taking into account the second letter (i.e. the digestive procedures). We performed different searches on the third letter (action), the fourth letter (access mode and/or technique) and also combined searches on both the third and fourth letters. The third letter (action) was Q “Examination” or H “Biopsy” for diagnostic procedures; And N “Destruction”, F “Exeresis”, A “Dilatation”, G “Ablation”, S “Hemostasis”, or J “Exsufflation” for therapeutic procedures. The fourth letter was E for “transorifice endoscopic access”.

For each kind of search we distinguished between therapeutic and diagnostic procedures, and measured precision and recall. Precision was defined as the number of relevant procedures retrieved by a search divided by the total number of procedures retrieved by that search, and recall was defined as the number of relevant procedures retrieved by a search divided by the total number of endoscopic procedures (which should have been retrieved).

	Diagnostic procedures				Therapeutic procedures				Diagnostic and therapeutic procedures			
	TP.	FP.	Recall	Precis.	TP.	FP.	Recall	Precis.	TP.	FP	Recall	Precis.
Manual search												
Manual selection of endoscopic procedures	35	-	-	-	86	-	-	-	121	-	-	-
Keyword search												
Procedures' labels containing "endoscopy"	13	1	0.37	0.92	27	0	0.31	1	40	1	0.33	1.0
Procedures' labels containing the suffix "-scopy"	35	4	1	0.90	85	120	0.99	0.41	120	124	0.99	0.49
Hierarchical search												
Procedures in hierarchical elements labels containing "endoscopy"	28	4	0.8	0.88	35	0	0.41	1	63	4	0.52	0.94
Procedures in hierarchical elements having at least 50% endoscopic procedures	28	4	0.8	0.88	68	14	0.79	0.82	96	18	0.79	0.84
Procedures in parent categories of both hierarchical elements selected above	35	127	1	0.22	75	372	0.87	0.17	110	499	0.91	0.18
Partial code search												
Procedures with "transorifice endoscopic access"	17	3	0.49	0.85	81	3	0.94	0.96	98	6	0.81	0.94
Procedures with action "Examination"	35	82	1	0.3	-	-	-	-	-	-	-	-
Procedures with action "Examination" or "Biopsy"	35	111	1	0.24	-	-	-	-	-	-	-	-
Procedures with "transorifice endoscopic access" and one action above	17	3	0.49	0.85	-	-	-	-	-	-	-	-
Procedures with action "Destruction", "Excision", "Dilation", "Ablation", "Hemostasis", "Exsufflation" or "Draining"	-	-	-	-	76	568	0.88	0.11	-	-	-	-
Procedures with "transorifice endoscopic access" and one action above	-	-	-	-	73	3	0.84	0.96	-	-	-	-

Table 5: Comparative table for several search strategies

Results

Table 5 shows results using the three different strategies and their variants with precision and recall. TP means true positive and FP means false positive. There are separate columns for diagnostic procedures, therapeutic procedures and both.

Full-text search: Recall was better using “scopy” but precision was higher using “endoscopy”.

Hierarchical tree search: Precision was high using only procedures in hierarchical levels containing “endoscopy” but recall was lower than in other strategies involving the CCAM hierarchy. For example recall was high when using parent categories of the selected branches but precision was very low. It was necessary to combine 22 different branches of Chapter 7 (2 for diagnostic procedures and 20 for therapeutic procedures). Among these 22 branches the label of the hierarchical level contained the strings “endoscopy” or “endoscopic” in 8 cases. The number of hierarchical levels that subsume these 22 branches was equal to 8.

Code search: The best results were obtained with the transorifice endoscopic access (precision 0.94 and recall 0.81). For diagnostic procedures recall was high using “examination” as an action whereas taking the “biopsy” into account was in relation with lower precision. Searching by code was in some cases inefficient when searching by access modes for procedures with multiple approaches, for example HHQJ002 - colonic echo-endoscopy without biopsy where letter J means “with ultrasound guidance”.

Discussion

From data to knowledge

Storing data in the computerized medical record using a controlled vocabulary should help to identify relevant information for clinical and medico economic studies [10]. The effort to improve quality of coding will facilitate the use of those data sets for such studies [11]. An appropriate knowledge representation is a requisite for an efficient retrieval of data and its transformation into new knowledge or information for policy making [12-13]. Medical procedures entered in the case mix database are a valuable tool for such studies.

Conclusion on these three search methods

In order to conduct efficient information searches it is necessary to combine several approaches, e.g. by keyword and hierarchy, and mix several hierarchical levels or several descriptors. These methods are not effective in the case of procedures

combining several actions or surgical approaches. In the case of hierarchical searches it is difficult to combine the content of 22 different branches. One convenient property of CCAM is that procedures that share common techniques such as endoscopes are usually grouped within a given anatomical part in the CCAM hierarchy. Using the CCAM codes is efficient but requires good knowledge of the CCAM in order to identify which descriptors are the more relevant. There is no method of selecting procedures by endoscopy “in one click”. Table 6 is a summary of the most important advantages and disadvantages of each method.

Search method	Advantages	Disadvantages
Full-text	The keyword search could be done on labels, notes or chapters headings.	Necessary to choose between precision and recall
Hierarchical	A good consideration of the anatomical location and the ability to combine it with text searching.	It is impossible to select the diagnostic and therapeutic procedures in the same query.
Code	This code search is well suited to multiple criteria search on “elementary procedures”, e.g. with a single action and / or a single access mode.	It fails on complex procedures with multiple actions / access mode.

Table 6: Summary of main advantages and disadvantages for each search method.

Limits

We implemented a single scenario on endoscopic procedures, worked on a single axis related to equipment (endoscopes), and used a single chapter (digestive procedures). The hierarchical organization and semi structured codes may have other properties in other scenarios for example procedures performed in medico technical units (radiology, operating room, interventional cardiology, functional explorations). This work may benefit from better techniques for full text search using natural language processing. Variations at the semantic level require semantic and lexical resources for synonyms.

Perspectives

In the next step we will apply the three search strategies to our hospital case mix database in order to measure the number of discharge datasets related to the selected CCAM codes and to assess the activity of the digestive endoscopy department.

When considering other multi axial terminologies results of the lexical and hierarchical searches may be compared but search by code is not relevant for terminologies that use non semantic Concept Identifiers. For example procedures in SNOMED

CT may be associated to several anatomies, actions and/or access modes according to its logical representation. Logical Search should be substituted to search by code for such terminologies.

We plan to implement a formal representation which allows finding the information in a more user friendly way: In this “Galen-like” representation, all procedures performed under endoscopy are labeled with the relationship *BY_MEANS_OF Endoscope*. Figure 1 is an example of a GALEN intermediate dissection that was written in order to represent a CCAM code in the GALEN-in-Use European project [3].

RUBRIC *"Suture d'une perforation de l'oesophage thoracique, par thoracoscopie, avec ou sans drainage externe de l'oesophage"*

```

MAIN suturing
    ACTS_ON breach
        HAS_LOCATION oesophagus
            IS_CONTAINED_IN
                thoracic cavity
                    BY_APPROACH_TECHNIQUE inspecting
                        ACTS_ON thoracic cavity
                            BY_MEANS_OF endoscope
                                WITH_OPTIONALLY draining
                                    ACTS_ON oesophagus
  
```

Figure 1: Example of a Galen rubric

We propose to develop tools and software for retrieving medical procedures in the case mix database. The search would thus benefit from all relations and descriptors proposed for the description of the CCAM in the GALEN project whether for anatomy, action, technique or surgical approach with more granular descriptors and possibility of combining several anatomies, actions, techniques or surgical approaches.

There is currently no available intelligent tool for retrieving CCAM procedures that fits all kinds of needs for medico economic analysis of hospital activity. None of the three search strategies is totally satisfying and this will lead to a new project for developing a new tool with better knowledge representation of medical procedures.

Acknowledgments

This work was partially granted by the InterSTIS project funded by the French National Research Agency (ANR-07-TECSAN-010).

References

1. Trombert-Paviot B, Rector A, Baud R, Zanstra P, Martin C, van der Haring E, Clavel L, Rodrigues JM. The development of CCAM: the new French coding system of clinical procedures. *HIM J.* 2003;31(1):1-11.
2. Rodrigues JM, Kumar A, Bousquet C, Trombert B. Standards and biomedical

- terminologies: the CEN TC 251 and ISO TC 215 categorial structures. A step towards increased interoperability. *Stud Health Technol Inform.* 2008;136:857-62.
3. Trombert-Paviot B, Rodrigues JM, Rogers JE, Baud R, van der Haring E, Rassinoux AM, Abrial V, Clavel L, Idir H. GALEN: a third generation terminology tool to support a multipurpose national coding system for surgical procedures. *Int J Med Inform.* 2000;58-59:71-85.
4. Bouchet C, Bodenreider O, Kohler F. Integration of the analytical and alphabetical ICD10 in a coding help system. Proposal of a theoretical model for the ICD representation. *Stud Health Technol Inform* 1998;52 Pt 1:176-9.
5. Franz P, Zaiss A, Schulz S, Hahn U, Klar R. Automated coding of diagnoses - three methods compared. *Proc AMIA Symp.* 2000:250-4.
6. Michel PA, Lovis C, Baud R. LUCID: a semi-automated ICD-9 encoding system. *Medinfo* 1995;8 Pt 2:1656.
7. Etude concernant l'assistance au codage et à la saisie des actes CCAM. Groupement pour la modernisation des systèmes d'information hospitalier. [Online] 09 2003. [Cited: 07 09, 2010.] http://www.gmsih.fr/fre/content/download/418/2719/file/121.Aide_CCAM-V0-79.pdf.
8. Téléchargement de la CCAM. AMELI - L'assurance maladie en ligne. [Online] [Cited: 07 09, 2010.] <http://www.ameli.fr/accueil-de-la-ccam/index.php>.
9. Guide de lecture et de codage. Classification commune des actes médicaux. Annexe I de l'arrêté du 27 février 2007. *Bulletin officiel* No 2007/3 bis. [Online] 2007. [Cited: 07 09, 2010.] <http://www.sante-sports.gouv.fr/IMG/pdf/bo0703.pdf>
10. Cimino JJ. From data to knowledge through concept-oriented terminologies: experience with the Medical Entities Dictionary. *J Am Med Inform Assoc* 2000;7(3):288-97.
11. Lorenzoni L, Da Cas R, Aparo UL. The quality of abstracting medical information from the medical record: the impact of training programmes. *Int J Qual Health Care* 1999;11(3):209-13.
12. Joubert M, Fieschi M, Robert JJ, Volot F, Fieschi D. UMLS-based conceptual queries to biomedical information databases: an overview of the project ARIANE. *Unified Medical Language System. J Am Med Inform Assoc* 1998;5(1):52-61.
13. Mendonça EA, Cimino JJ, Johnson SB, Seol YH. Accessing heterogeneous sources of evidence to answer clinical questions. *J Biomed Inform* 2001;34(2):85-98.