Building an Ontology of Hypertension Management

Olivier Steichen, Christel Daniel-Le Bozec, Marie-Christine Jaulent, and Jean Charlet

INSERM, UMR_S 872, Eq. 20, F-75006 Paris, France
ost@club-internet.fr

Abstract. The analysis of customized decisions during hypertension management in a specialized unit requires a detailed representation of clinical cases. We are building a specific ontology to code medical records and process them with computerized tools. Relevant concepts to describe and justify medical decisions are extracted from three sources: (i) Clinical guidelines; (ii) Items of the semi-structured medical record form used in the clinical unit; (iii) Free-text answers from 5,000 completed record forms. Combining terminological sources is mandatory to cover the whole spectrum of possible justifications for clinical decisions, including contextual specificities and patients' particulars.

Keywords: Ontologies, Computerized Medical Records, Guidelines.

1 Introduction

1.1 Qualitative Practice Assessment

Clinical guidelines are often taken as references for practice assessment. However, clinical decisions sometimes have to be customized, beyond or against guidelines, with regard to the characteristics of each patient and specific clinical circumstances. However, even customized decisions must remain justified. They can be assessed with qualitative methods that rely on detailed and comparative analysis of individual cases [1]. Computerized tools are mandatory to achieve such studies on a large scale. They can help clustering similar cases according to clinical characteristics and they can ease the comparative study of medical decisions, provided that medical records are formally represented with enough details for the intended level of analysis.

1.2 Medical Ontologies Modeling

Ontologies are considered as the richest and most formalized terminological resources to share and process the semantic content of electronic health records. As such, they can meet our requirements for medical records representation. Because each ontology has a specific conceptual scope, granularity level and structure, it usually can't be reused for purposes it wasn't primarily intended

R. Bellazzi, A. Abu-Hanna, and J. Hunter (Eds.): AIME 2007, LNAI 4594, pp. 292–296, 2007.

for [2]. In order to assess customized clinical decisions in a hypertension unit, we are building a specific ontology of hypertension management to represent the content of medical records. This paper describes our practical ontological modeling process.

Two broad approaches compete to elicit concepts to be included in an ontology. Top-down approaches rely on prior conceptualizations made by field experts (in clinical medicine: structured data entry forms, clinical guidelines, academic books, etc.). In bottom-up approaches, concepts are extracted from the experts' textual production during actual practice (in clinical medicine: medical records, discharge summaries, mails to colleagues, etc.). We followed a custom composite approach, combining top-down and bottom-up steps.

2 Material and Methods

We used three sources to find the concepts used to describe and justify medical decisions during hypertension management: (i) Items of the semi-structured medical record form used in the hypertension unit, representing the generic determinants of practice in specialized settings; (ii) A corpus of clinical guidelines, representing the generic determinants of practice in family medicine; (iii) A corpus of free-text comments from filled record forms, allowing physicians to consign the unforeseen determinants of specialized practice and the justifications of difficult decisions.

A computerized semi-structured medical record is used for 30 years in the hypertension unit and counts 176 questions [3]. The list of clinical items found in the form (headings, questions and predefined answers of checklist questions) has been generated by a SQL query on the hospital database. After anonymization, the free-text answers from the 5,109 clinical record forms filled in year 2005 resulted in a 350,000 words corpus in French. Eight hypertension guidelines, published from 1999 to 2005, were assembled to constitute a 56,000 words corpus in English.

The clinical record form and the guidelines were the starting points of our first two ontological modeling steps, both top-down. They provided core concepts for hypertension management. The free-text answers from completed forms fed a bottom-up modeling step, intended to enrich this ontological core.

The natural language processing tools SYNTEX and UPÉRY were used to extract term candidates (terms likely to represent pertinent concepts in the field) from guidelines and free text answers [4]. Ontological modeling is currently carried out, concept by concept, in the Differential Ontology Editor (DOE) [5]. Concepts are manually linked with SNOMED-CT concepts (January 2006 version), accessed through the Virginia Tech SNOMED-CT Browser [6]. This mapping of our domain ontology with a reference ontology gives an opportunity to evaluate the domain coverage of SNOMED-CT and the adequacy of its organization with regard to our intended use.

3 Results

Concepts from the record form items and from guidelines have been completely extracted. By now, only the most frequent concepts in the free-text comments – occurring at least 50 times in the corpus (about 1% of the records) – have been extracted. These concepts mostly relate to some generic aspects of hypertension management, as do concepts from the record form items and from guidelines. Therefore, we presumably have identified most of the core concepts used for routine hypertension management. We are currently structuring this ontological core. Only concepts coming from the record form items have been linked to SNOMED-CT at the present time.

3.1 Top-Down Steps: Considering Experts' Conceptualizations of Their Field

The 176 questions and 177 predefined answers of checklist questions in the record form provided 243 clinical concepts. The same concept can underlie several items, for example a date question ("date of the last myocardial infarction?") and a predefined answer of a checklist question ("heart history?" ☒ myocardial infarction). The relevance of all these concepts is assumed, given the clinical appropriateness of the form as a result of step-by-step improvements over 30 years. We found a matching SNOMED-CT term for 212 of these concepts. Guidelines analysis uncovered 258 clinical concepts: 163 concepts already found in the form items and 95 additional concepts.

The concepts found in the form items but not in the guidelines were either too specialized (like renal infarction as a cause of hypertension), too fine-grained (like glomerulopathy as a specific type of kidney disease) or related to other cardiovascular risk factors management (like diabetes or dyslipidemia). Two examples of additional concepts found in guidelines but not in the form items are "sleep apnea" as a cause of hypertension and "dementia" as a consequence of hypertension.

3.2 Bottom-Up Steps: Considering Free-Text Answers in Medical Records

The analysis of term candidates in free-text comments revealed 233 frequent concepts (more than 50 occurrences). Among them, 162 had already been found in the record items or in the guidelines and 71 were completely new.

Concepts found in guidelines and actually used by physicians in free-text answers are undoubtedly valuable for the management of patients in the specialized unit. For example, free-text comments count 89 occurrences of "sleep apnea" or of semantically related terms, like "snoring". On the other hand, no occurrence of "dementia" was found in free-text comments. According to the physicians, dementia is not an issue faced in the hypertension unit. Patients are simply not referred by their general practitioner if they suffer from disabling cognitive impairment.

The analysis of free-text comments also identified concepts missing in guidelines but regularly used by physicians to state or justify their decisions. Most of these concepts are very specialized and refer to the specific recruitment of the unit (high prevalence of secondary and/or complicated hypertension), whereas guidelines are mainly intended for the management of primary and uncomplicated hypertension in general practice. For example, the concept of renin – aldosterone dissociation, related to specific causes of hypertension, has 80 occurrences in free-text comments and is an important determinant of patients' management, with respect to local decision rules or habits.

4 Discussion – Conclusion

4.1 The Need for a Customized Domain Ontology

An ontology intended to support a specific task in a specialized domain requires a definite expressive power (scope and granularity level) and processing potential (structure). The specific shortcomings hindering a direct reuse of SNOMED-CT for our application were highlighted during concept mapping. Some SNOMED-CT terms are ambiguous and unequivocal mapping was sometimes difficult. For example, the concept of glomerulopathy, found as a kidney disorder in the record form, could be mapped with two brother SNOMED-CT terms, which are clearly synonyms for a clinician: "renal glomerular disease" and "glomerular disease". As expected, the scope of SNOMED-CT is not broad enough. Thirty-one of the 243 fundamental clinical concepts coming from the form items were lacking. For example, there is no term in SNOMED-CT referring to monogenic hypertension or drug induced hypertension. Finally, the structure of SNOMED-CT is presumably inadequate for our purposes, because it doesn't reflect the way clinicians picture the field of hypertension medicine. For example, the structuring concept of target organ damage, which determines many management decisions, is not found in SNOMED-CT.

4.2 The Need for a Customized Modeling Approach

Our ontological modeling approach relates with former bottom-up achievements relying on large corpora of texts produced during clinical practice [4,7]. This framework ensures that the ontology comprises the concepts actually used by physicians. Nonetheless, we customized this approach for our particular needs: (i) Corpora used in former works were collections of discharge summaries whereas we worked with medical records; (ii) We added top-down steps, exploiting pre-existent experts' conceptualizations of the field.

Knowing that a concept is found among the record form items or in guidelines is a strong hint in favor of its contribution to standardized hypertension management, following explicit or implicit rules. We also considered frequently occurring concepts in free-text comments, which must be related to some aspect of routine hypertension management. The remaining, less frequently occurring concepts in free-text comments are likely to play a role in the individualization of management decisions.

4.3 Future Work

We are currently organizing our ontological core in a strictly taxonomic hierarchy. We will then complete the terminological analysis of free-text comments, in order to integrate the less frequent concepts in the ontology. After a strong initial focus on concepts not related to practice individualization, it should be easier to discern and incorporate concepts more directly related to our purpose.

We will then use the ontology to represent clinical cases managed in the hypertension unit. Free-text answers can not be automatically represented with the concepts they issued. The association of matching concepts to each case can only be automated for the structured part of medical records. Indeed, the form items are unequivocally linked with the concepts they gave birth to. Once this partial representation achieved, it will be possible (i) to identify cases managed outside guidelines' recommendations and (ii) to cluster similar such cases thanks to semantic similarity measures.

Within cluster comparisons, i. e. qualitative analysis of similar cases, will lead to uncover further pertinent characteristics of cases and their ontological representation will be manually complemented. After a first loop, it will be possible to compute semantic similarity anew on the extended ontological representations. Finer clusters of similar cases will ensue, allowing a finer qualitative analysis. The process may be looped as many times as useful and possible.

Acknowledgements. Didier Bourigault provided the statistical results of corpus analyses with SYNTEX and UPÉRY. This work benefits from a grant of the AP-HP and from a subsidy of the French Society for Hypertension.

References

1. Green, J., Britten, N.: Qualitative Research and Evidence Based Medicine. BMJ 316, 1230–1232 (1998)
2. Coiera, E.: Medical Informatics. BMJ 310, 1381–1387 (1995)
3. Degoulet, P., Chatellier, G., Devriès, C., Lavril, M., Ménard, J.: Computer-Assisted Techniques for Evaluation and Treatment of Hypertensive Patients. Am J Hypertens. 3, 156–163 (1990)
4. Charlet, J., Bachimont, B., Jaulent, M.: Building Medical Ontologies by Terminology Extraction from Texts: An Experiment for the Intensive Care Units. Comput Biol Med. 36, 857–870 (2006)
5. Troncy, R.: Differential Ontology Editor (Last accessed on the 2006/09/09) `http://homepages.cwi.nl/~troncy/DOE/`
6. Virginia-Maryland Regional College of Veterinary Medicine: SNOMED-CT Browser (Last accessed on the 2007/01/30) `http://snomed.vetmed.vt.edu/sct/menu.cfm`
7. Baneyx, A., Charlet, J., Jaulent, M.: Building An Ontology Of Pulmonary Diseases With Natural Language Processing Tools Using Textual Corpora. Int J Med Inform. 76, 208–215 (2006)