Mapping Cancer Patients' Symptoms to UMLS Concepts

Laura Slaughter, Ph.D[1]., Cornelia Ruland, RN, Ph.D[1,2], Ann Kristin Rotegård, RN,MS[1]
[1]Center for Shared Decision Making and Nursing Research, Rikshospitalet, Oslo, Norway
[2]Department of Medical Informatics, Columbia University, New York, NY
{laura.slaughter|cornelia.ruland|ann.kristin.rotegard}@rikshospitalet.no

A prerequisite for patient-friendly personal health records (PHR) is their ability to allow seamless integration of patient terminology with professional terminologies. In this work, we mapped a set of symptoms/ problems from the self-assessment component of a cancer patient support system to concepts in the Unified Medical Language System (UMLS) Metathesaurus. Our objective was to learn how the UMLS can be used as a tool to connect patient terminology with professional vocabularies. The mapping to UMLS was done with the help of ten expert cancer nurses who evaluated concepts, their synonyms and placement in the source vocabulary hierarchical structure. The UMLS concepts were also compared with terms and phrases found in patient medical records that addressed the same set of symptoms. In this study we observed several problems related to the use of the UMLS Metathesaurus as a tool to connect from patient-level expressions to professional-level classification systems. More work is needed to increase interoperability between layperson health applications and clinical systems.

INTRODUCTION

Recent initiatives in the development of personal health records (PHR) emphasize the need for suitable methods that support seamless information sharing between clinicians and patients. Systems that are used by patients (such as for managing a chronic illness or for providing decision support) must contain an interface between the world view, language used, and conceptual organization of the patient and the available professional-level medical classification. This connection is crucial for numerous uses such as mapping to the electronic health record (EHR), to suggest potential interventions, and for use in health professionals' search for literature. Although there is research concerning the construction of consumer health vocabularies[1,2,3], there is a need for studies that investigate how a system of patient-level health expressions can be modelled and aligned with professional-level medical and health classification systems and, in the future, formal ontology. This approach offers 1) greater potential for interoperability between disparate systems and 2) added support for use of semantic-based functions (e.g. semantic search algorithms).

Two recent proposals have been put forth to create patient-level knowledge resources. Soergel et al.[4] observed ways in which patient terminology and relationships expressed between the terms lack agreement with biomedical knowledge. They proposed an "interpretive layer" to "assist consumers in identifying terms to describe their needs, finding, and understanding relevant information." They state that such a system needs two components: an educational consumer health ontology and representations of lay explanatory models. Smith and Fellbaum[5] describe a project to construct a lexical database "consisting of medically relevant terms used by and intelligible to non-experts" called Medical WordNet (MWN). In addition, they plan on creating two sentential subcorpora called Medical FactNet (MFN) and Medical BeliefNet (MBN). The three "Nets" form a foundation for supporting machine translation, question-answer systems, and text summarization. Their work to construct a fact network and a belief network in tandem allows "systematic assessment of the disparity between lay beliefs and the corresponding expert medical knowledge."

Our research focuses on connecting the patient view (e.g. such as those modelled in a MWN) with the expert biomedical domain and is therefore supporting the construction of lay health knowledge-based systems. This paper describes the mapping of patient symptoms from a support system for cancer patients to Unified Medical Language System (UMLS) Metathesaurus concepts. With the help of expert cancer nurses, we manually mapped words and phrases describing patient symptoms to UMLS concepts and evaluated their connection to biomedical knowledge structures. Furthermore, we investigated how patients' symptom /problem terms corresponded to nurses' professional terms describing the same symptoms in the EHR. By investigating the feasibility of representing and integrating lay language and professional practice terms with UMLS concepts this study addresses an important aspect in the development of the PHR.

METHODS

The symptoms we focus on in this study are a part of a support system for cancer patients called CHOICE[6]. This system allows patients to 1) perform symptom self-assessments (to facilitate patient-centred care), 2) obtain trustable information tailored to their needs, and 3) to support treatment/intervention decisions. There are two versions of the system, one in Norwegian (Norsk bokmål) and the other in English. The symptom list was developed for CHOICE based on a critical review of evidence-based literature and focus groups with cancer experts to identify the problems, specific symptoms, and functional limitations encountered by cancer patients. These symptoms were then "translated" into lay language through patient focus groups and a subsequent pilot study with 15 cancer patients[7]. The CHOICE system was recently tested in a randomized clinical trial (RCT), yielding 1540 assessments. We selected the ten symptoms that were most frequently selected by patients in the RCT for mapping to the UMLS.

1. Weakness *(slapp, utmattet)*
2. Lack of energy *(manglende energi/tiltaksløs)*
3. Much slower than usual in things I do *(bruker lengre tid enn vanlig på det jeg skal gjøre)*
4. Difficulty sleeping through the night *(vanskeligheter med å sove gjennom natten)*
5. Problems getting enough food/nutrition *(vanskeligheter med å få i meg mat)*
6. Lack of appetite *(manglende matlyst)*
7. Nausea *(kvalme)*
8. Taste/smell changes *(lukt eller smaksendringer)*
9. Mouth sores/ pain in mouth and throat *(sårhet eller smerter i munn og svelg)*
10. Mood swings *(humørsvingninger)*

The ten CHOICE symptoms mapped in this study.

There were two goals for this study: the first was to evaluate the UMLS concepts that mapped to the patient symptoms, and the second was to compare the symptoms and their mapped concepts with the terminology used by nurses in the patient charts.

Mapping & Evaluation of UMLS Concepts

For each symptom, we performed a lookup using the UMLS Knowledge Source Server. This tool locates terms in the UMLS, based on lexical similarity, that are either an exact match or a variation taking into account word stemming, inflection, inversion, and alphabetic case. In the CHOICE system, some symptoms are expressed as short, non-sentential phrases. We tried manual variations of these expressions to find concepts that would potentially map (e.g. by removing modifier "through the night" from *difficulty sleeping through the night*). We also used both the English terms and the Norwegian literal translation to English for mapping. For example, the

English symptom *weakness*, is *slapp/utmattet* in the Norwegian version which translates directly to *limp,listless/exhausted,worn out.*

In the UMLS, synonymous terms taken from a variety of vocabulary sources form a cluster which is called a *concept*. These concepts are classes that represent a given meaning. For example, the meaning "dysgeusia" is represented by the set of synonymous terms {dysgeusia, parageusia, disordered taste, gustatory alteration, taste abnormality} and forms the UMLS concept *C0013378*. The concept *C0013378* is assigned the preferred string "dysgeusia" in the system. The vocabulary sources that provide the synonymous terms form an ancestor hierarchy superordinate to the concept which can also have children (subordinate) concepts. These vocabularies also provide definitions and "other related and possibly synonymous" terms.

We asked ten cancer nurses to contribute their expert opinion in the evaluation of the mapped UMLS concepts. All the nurses who helped with the evaluation tasks were employed in a cancer ward and regularly cared for cancer patients. Some of the nurses were also part of a nursing informatics professional development course. The nurses were native Norwegian speakers and their skills in speaking, using, and understanding the English language ranged from proficient to fluent. All evaluation sessions were conducted in English. We used an English-Norwegian dictionary, a medical dictionary (in English), and also the UMLS Knowledge Source Server (KSS) as reference materials. The tasks that the nurses were asked to perform were as follows:

Task 1- Decide whether the UMLS concept that was mapped using the KSS is an appropriate match to the CHOICE symptom. Nurses did this evaluation by looking at the UMLS preferred string as well as any available UMLS definitions. An "appropriate match" is a concept that can be used to express the symptom.

Task 2- Evaluate the set of synonymous terms that form the concept and, in addition, the UMLS terms "other related and possibly synonymous terms." The nurses used these to create a subset of synonymous terms that describe the symptom as it is experienced by cancer patients. In some cases, the nurses commented on whether or not the patient might say that term, or whether the term is usually used by physicians or nurses as professional language. In addition, we asked nurses to provide the Norwegian equivalent terms connected to the concept.

Task 3- Evaluate the hierarchical structure (ancestors and children) from each contributing vocabulary source in order to a) locate potential parent or child terms that map to the symptom, and also b) comment on the classification structure.

Nurses were informed prior to beginning the sessions that the aim of this project was to map CHOICE symptoms to professional-level vocabulary for use in a patient-oriented care plan, and to help with retrieval of information based on the symptoms selected by patients. Together with the first author, we worked as a team in the evaluation process. In most sessions, two nurses worked together to complete the tasks for one or two symptoms. Because of scheduling conflicts, in two sessions, only one nurse worked with the first author on the tasks. Some symptoms mappings were evaluated by two separate pairs of nurses in order to compare comments and results.

Comparison with Terms in Patient Charts
In addition to the evaluation of mapped UMLS concepts, the third author, who is an expert cancer nurse, identified expressions of symptoms in 16 cancer patients' charts. A systematic chart abstraction was done in order to extract terms and phrases from the nursing documentation that mapped to the 10 selected symptoms from the CHOICE system. The nursing documentation consists of Norwegian natural language organised in narrative text as well as care plans (these have not been structured according to standard nursing classification models). Each word or phrase that was synonymous to a CHOICE symptom was then compared with the English terms from the UMLS synonym list.

RESULTS

Synonymous Terms for each Symptom
This work resulted in sets of synonymous terms that represent the extension of each patient symptom. To illustrate, Table 1 shows the UMLS mappings for two of the symptoms, their English synonyms that form extension sets, and their Norwegian language equivalents. We have also included information concerning aspects related to the use of each term (e.g. whether the term is used by patients, used by professionals, and/or identified in patient charts). We also show the connections to the terminology used by nurses within patient charts. The Norwegian terms have English equivalents within the UMLS in most cases. In the nurses' natural language terminology, the symptoms are often connected to where or how the symptom is expressed by the patient. The nurses also use synonyms in the written nursing texts that they did not mention in the formal discussions of the UMLS concepts. Through the chart abstraction, we were able to connect nursing note terminology to UMLS concepts.

Our primary results focus on the connection between patient expressions and UMLS concepts. Through the process of mapping these ten patient symptoms, we made a number of important observations concerning use of the UMLS Metathesaurus as a tool for connecting our CHOICE symptoms to professional vocabularies:

No UMLS concept available: Some symptoms in the CHOICE system are functional problems that patients experience, rather than true symptoms. For example, *much slower than usual in the things I do* is a self-evaluation of one's ability to perform normally during their daily routine. Nurses supported the need to list this separately in the CHOICE system, distinct from *lack of energy*. It is important to address this as patients try to resume a normal life following treatment. Ignoring this problem, when the patient is bothered with it, can lead to increased frustration and depression. However, we were unsure of how to map this to a UMLS concept. According to nurses' expert opinion, the closest match might be *fatigue (C0015672)*. Another possibility might be to map this to *performance of usual routine (C0517352)* which is a concept from the Nursing Outcomes Classification. It might also be related to the concept *Endurance (C0518031)*. In the charts, the nurses are documenting this symptom as "have to take breaks" and "gets tired rapidly."

A group of concepts can map to one symptom: We have found that, in some cases, multiple concepts must be mapped to a symptom in order to represent the patient experience more completely. For example, the symptom listed in CHOICE, *mouth sores, soreness or pain in mouth and throat (sårhet eller smerter i munn og svelg)* can be mapped to: *swelling & soreness of mouth (C0857227), sore mouth (C0423615), sore gums (C0239739), tender mouth (C0877460), stinging mouth (C0877461), sore roof of mouth (C0542218),* and *painful mouth (C0221776)* (although two nurses were hesitant to include *painful mouth* in this set, since they felt the concept is too broad). Nurses rejected the concept *sore throat (C0242429)* because this is a symptom usually associated with a cold, which is different from the cancer patients' experiences. They also rejected *soreness corner mouth (C0877484)* because of lack of understanding of what is meant by this concept. Phrases found in the charts show us that nurses document this symptom using language we didn't consider when we were looking up concepts for the nurses' evaluation. Nurses write, for instance, "discomfort in mouth/throat", "irritation of throat", and "mucosa pain." Some of these phrases can be found in the UMLS but conflicts arise between the language used in the charts and the concepts chosen by the nurses, for example, the phrase written in the charts "discomfort in throat" is a synonymous term

for the concept *sore throat* which was rejected by the nurses.

Concept class contains errors in the set of synonymous terms: In some cases, the nurses reported that some of the terms in a concepts' set of synonyms are separate concepts (are not really synonyms). In the case of *dysosmia (C0235287)*, for example, there are many synonymous terms that seem "right" (e.g. *distorted sense of smell*, *perversion of the sense of smell*) and others that do not seem to belong to this set, such as *parosmia* and *cacosmia*. Unfortunately, the definition of *dysosmia* in the UMLS, taken from the Home Health Care Classification, "diminished ability to smell" is not consistent with standard medical dictionaries. Dysosmia is a "distortion of normal smell perception"[8]. In the concept synonym set, the terms *cacosmia* is "subjective perception of a disagreeable odor," and *parosmia* is "a false sense of odors or perception of those that do not exist." Neither of these was considered a synonym, although they are related terms, of *dysosmia*. The nurses modified the set of synonyms for *dysosmia* removing all that were inappropriate.

Inconsistency hinders mapping: The nurses described the patients' experience of the symptom *smell and taste changes* as "overly sensitive to smells, and that patients feel smells are much stronger." At first, we explored the concept *dysosmia (C0235287)*. However, through exploration of "Related and Possibly Synonymous Terms", nurses identified what they considered to be a very good concept match to the symptom, having the preferred string *Sensitive to smells (C0234259)*. On closer inspection, we discovered that *Sensitive to smells (C0234259)* probably means *hyperosmia* "an abnormal sensitivity to odors." The inconsistency between concepts related to smell (e.g. *hyperosmia (C0234259)* and *dysosmia (C0235287)*) caused difficulties in finding the appropriate match.

Ancestor Hierarchy
Nurses' evaluation of the ancestor hierarchy resulted in one additional important observation.

Subgraphs are necessary: When nurses evaluated the ancestor hierarchies for each concept, they commented on the structure and at times, suggested removal of certain vocabulary sources. The nurses discussed that the mapped UMLS concepts can sometimes carry additional meaning inherited from the parent terms that are not related to the cancer patients' experiences. This is seen in the mapping of the symptom *lack of appetite (Norwegian term: manglende matlyst)* to the concept *anorexia*

(C0003123). Some of the nurses interviewed said that the concept *anorexia* is "associated with mental disorder which is not what is happening with these patients" while others said that "although it isn't used, it is still the correct professional term." Nurses suggested removal of some hierarchies such as *Psychiatry and Psychology (MeSH Category)* → *Mental Disorders* → *Eating Disorders* → *Anorexia*. Looking at nurses' terminology within the patient charts demonstrates that nurses mainly document that appetite is poor, decreased, or absent. A mapping to *decrease in appetite* (C0232462) instead of anorexia will not solve the problem because this concept is also linked to ancestors of *eating disorders*.

DISCUSSION

We found that it is important to look closely at each UMLS concept mapped (e.g. check all synonymous terms) and consider clusters of symptom concepts. Looking from the nursing notes to UMLS concepts is very helpful and provides another source of data to assess the mapped concepts. Our results support the findings of Travers et al.[9] that nurses' natural language in patient's records can be connected with UMLS but, in general, nurses use more specific terminology than the UMLS concepts.

Within our sample of 10 patient expressions, we have shown that using the UMLS requires a great deal of manual analysis to locate the concept(s) that best reflect each symptom. Connecting symptoms experienced by cancer patients to the UMLS concepts is not a straightforward task even with the help of experts familiar with both patient and medical viewpoints. If the UMLS will be used as a means to connect patient (e.g. in CHOICE and PHR) and professional systems (e.g. in EHR), then it must contain terms for the important signs and symptoms as they are experienced by patients.

Our future goals are to work on methods to integrate CHOICE symptoms into the electronic nursing care plan. The transfer/integration of information to and from the PHR is important because this can lead to 1) a more complete and correct picture of the patient's situation, 2) more effective interventions, and 3) higher patient satisfaction - which can lead to shorter hospital stays and less complications.

Table 1: UMLS concept mappings for two CHOICE symptoms.

UMLS concept	English Synonym Subset	Norwegian Terms in Charts
Symptom: mouth sores (sårhet eller smerter i munn og svelg)		
Sore mouth (C0423615)	Sore mouth [1], Sore mouth (finding), Sore mouth – symptom	Sår munn [1], Munnsår/munnsårhet [3], Mucocitt [2,3]
Swelling & soreness in mouth (C0857227)	Swelling & soreness in mouth	Hoven og sår i munnen [3], Sårhet i munn /munnhulen[3], Sår på tungen[3], Hoven tunge[4], Hovne kinn[4]
Sore gums (C0239739)	Sore gums, Gingival tenderness, Sore gums-symptom, Soreness gum	None
Painful mouth (C0221776)	Painful mouth, Oral cavity pain, Oral pain, Painful mouth (finding), Pain mouth	Smerter i munnen[3], Vond tunge[4], Vondt i kjeven[3], Vondt å svelge[4], Smerter i halsen [4]
Tender mouth (C0877460)	Tender mouth, Mouth tenderness	Ømhet i kinn/tannkjøtt[4]
Stinging mouth (C0877461)	Stinging mouth	Sårhet i hals[4,] Sårhet i munnhule[4], Sårhet i svelg[4]
Sore roof of mouth (C0542218)	Sore roof of mouth	None
Not evaluated, from charts only	Discomfort in mouth/throat, Irritation of throat, Mucosa pain, Stinging mucous membrane, Sores in mucous membrane, Encrustation on tongue, Bleeding in mouth and throat, Blisters in mouth/on lips	Ubehag i munn /hals[3], Irritasjon i halsen[3], Mucositt smerte[3],Såre slimhinner[3], Munnslimhinner fulle av sår[3], Skorper på tunga[3], Småblødninger i munn og svelg[3], Blemmer i munnen/leppene[3]
Symptom: lack of appetite (manglende matlyst)		
Anorexia (C0003123)	Anorexia [2], [D] Anorexia NOS [2], [D] Anorexia NOS (context-dependent category) [2], [D] Appetite loss [2], [D] Appetite loss (context-dependent category) [2],Anorexic [2] Appetite impaired, Appetite lost, Inappetance, Loss of appetite [1], Loss of appetite (finding)	None
	Appetite absent, Lack of appetite [1],No appetite	Manglende appetitt [3]
Anorexia symptom (C0426579)	Anorexia symptom	None
Loss of appetite - symptom (C0426583)	Loss of appetite – symptom	None
Appetite loss - anorexia (C0426580)	Appetite loss - anorexia	None
No interest in food (C0426581)	No interest in food	None
decrease in appetite (C0232462)	Reduced appetite	Redusert matlyst[3], Nedsatt matlyst /appetitt[3]
	Poor appetite	Dårlig matlyst[3], ikke så god matlyst[3], liten matlyst[3], spiser lite[3]

Codes: 1 nurses report the term is patient-friendly   2 nurses report the term is not used by patients (professional)
3 term used by nurses in the charts   4 term used by nurses in the charts, but more specific than the UMLS concept

REFERENCES

[1] Zielstorff RD. Controlled vocabularies for consumer health. J Biomed Inform 2003; 36(4-5):326-333.

[2] Tse T, Soergel D. Exploring medical expressions used by consumers and the media: an emerging view of consumer health vocabularies. AMIA Annu Symp Proc 2003;674-678.

[3] Zeng Q, Kogan S, Ash N, Greenes RA, Boxwala AA. Characteristics of consumer terminology for health information retrieval. Methods Inf Med 2002; 41(4):289-298.

[4] Soergel D, Tse T, Slaughter L. Helping healthcare consumers understand: an "interpretive layer" for finding and making sense of medical information. Medinfo 2004; 2004:931-935.

[5] Smith B, Fellbaum C. Medical WordNet: A new methodology for the construction and validation of information resources for consumer health. Proc Coling: The 20th International Conference on Computational Linguistics 2004; Geneva 23-27 August.

[6] Ruland CM, Andersen R. Designing WEB- CHOICE. Medinfo 2004; 2004(CD):1840.

[7] Ruland CM, White T, Stevens M, Fanciullo G, Khilani SM. Effects of a computerized system to support shared decision making in symptom management of cancer patients: preliminary results. J Am Med Inform Assoc 2003; 10(6):573-579.

[8] Donald Venes Ed. Taber's Cyclopedic Medical Dictionary. Philadelphia: F.A. Davis Company, 2001.

[9] Travers DA, Haas SW. Using nurses' natural language entries to build a concept-oriented terminology for patients' chief complaints in the emergency department. J Biomed Inform 2003; 36(4-5):260-270.