Identifying and Characterizing a "Consumer Medical Vocabulary"

Tony Tse, Doctoral Candidate
College of Information Studies, University of Maryland
tsetony@oriole.umd.edu
Draft Dissertation Topic

Problem: To identify and analyze similarities and differences between professional, intermediary ("media"), and lay medical vocabularies (e.g., myocardial infarction vs. heart attack).

Questions

- What are *consumer medical vocabularies* (CMVs)?

- What are the characteristics (form and meaning) of terms in a CMV?

- What are the differences among CMVs, *media medical vocabularies* (MMVs), and *professional medical vocabularies* (PMVs)?

- How do health consumers categorize medical concepts?

Hypotheses

1. CMV terms will have greater overlap with PMVs in particular areas (e.g., signs and symptoms – fever, cough), but not in others (e.g., disease conditions – stomach ache).

2. Concepts represented by a CMV are a subset of those in PMVs.

3. Terms in a CMV will be generally broader than terms in PMVs.

4. MMVs consist of medical terms and concepts from both CMVs and PMVs.

Importance of the Research

As consumer demand for health information increases, improved methods for rapid and accurate information access by health care consumers are needed. It is well known that health professionals and laypersons may use different terminology, but the extent of the difference is not clear. Currently, online consumer health resources use PMV-based (e.g., Medical Subject Headings or MeSH) or "locally derived" MMVs to categorize medical information. An understanding of CMVs will provide insights into how health care consumers conceptualize health concepts and could lead to the development of consumer-friendly navigational aids in system interfaces to help consumers find information by leveraging their existing conceptual models. In addition, systems could be developed to help educate consumers based on their needs and to address common misconceptions about specific medical concepts.

Prior Research

Little work in the area of health consumer vocabularies has been reported in the literature. The few medical vocabularies that have been designed specifically for health care consumers (e.g., Planetree Classification Scheme and WellMed's Consumer Health Terminology (CHT)

Thesaurus) were created by "intermediaries" and represent MMVs. No studies have been found exploring the nature of CMVs relative to intermediary vocabularies (MMVs) – a mediating factor between consumers and professionals – and professional vocabularies (PMVs). Relevant findings from information science, medical informatics, cognitive psychology, and linguistics will be integrated into a conceptual framework.

Approach/Methodology

1. **Extracting Terms.** Terms used by health care consumers will be harvested from email messages, queries at consumer health Web sites, and online patient postings. Terms used by "third-party intermediaries" such as the mass media and Web-based resources (collectively referred to as "media" medical vocabularies or MMVs), intended to bridge the knowledge gap between health care professionals and health care consumers, will be extracted from various printed and electronic publications. Coders will select health-related terms from source material based on a small set of empirically derived criteria. The sources will represent a broad range of medical topics and, for MMV sources, factors such as intended purpose (advertisement versus information) will be distinguished.

2. **Characterizing the Vocabularies.** Lexical analysis of CMV and MMV terms will provide information about term complexity and degree of overlap in the use of terms among the domains. CMV and MMV terms will be mapped to high-level semantic types in the UMLS (e.g., Diseases or Syndromes and Signs & Symptoms). Terms for which there are no exact matches will be analyzed using "fuzzy matching" algorithms and, ultimately, manual review.

3. **Validating the Vocabularies.** The collected terms will be compared with terms found in queries from two health consumer Web sites (MEDLINE*plus* and *ClinicalTrials.gov*). The purpose is to ensure that actual consumer health needs are reflected in the derived vocabulary.

4. **Comparison with PMVs.** The derived MMV and CMV will be compared to the PMVs in the UMLS Metathesaurus. Terms existing only in the CMV and the MMV will be analyzed in detail. The degree and nature of any overlap with PMV concepts will also be investigated.

5. **Health Care Consumer Review.** Cognitive knowledge structures and mental associations used by consumers in the medical domain will be elicited by a variety of methods, such as concept mapping and word association tasks, selected source materials, and subsets of the collected terms. Such a preliminary study involving health care consumers will also help to assess the credibility and external validity of the derived CMV.

Questions

1. What criteria should be used to include a term in the Professional Medical Vocabulary (PMV) and in the Consumer Medical Vocabulary (CMV)

2. How can consumer "health-related" terms be identified (criteria)? Are automated methods possible without a "core" (seed) set of training documents and/or domain-specific lexicon?

3. Is the medical vocabulary used by the media/government/health organizations separate and distinct (apart from the CMV and PMV) or does simply "bridge" CMV and PMV (using terms from both)?