

Using Concept Markers to Find Genetics Content in a Medical School Curriculum

Jeffrey D. Smithers - Joshua C. Denny - Anderson Spickard III, MD, MS - Randolph A. Miller, MD

DEPARTMENT OF BIOMEDICAL INFORMATICS, VANDERBILT UNIVERSITY MEDICAL CENTER

PURPOSE

We sought to determine where genetics content was taught in our medical school curriculum.

BACKGROUND

At a recent Vanderbilt University School of Medicine (VUSM) curriculum retreat, a faculty member asked, "How well and where are we covering genetics in our curriculum?" This followed a proposal to develop a free-standing genetics course. At the time this question was posed, we were completing development of KnowledgeMap (KM) concept indexer. KM identifies UMLS concepts in free-text medical documents (such as curriculum handouts) and was designed for concept-based searching of online medical school curriculum. We have nearly 80% of VUMC lecture content of the first two years available in digital form. We used a subset of these documents and two possible sources of markers for genetics content to evaluate the utility of KM to address the aforementioned genetics question.

METHODS

The approach used was to independently identify "genetics concepts" and determine in which VUSM curricular documents they occurred. We began by identifying two sources of genetics terms to evaluate as markers for genetics content. The first source was the NIH National Human Genome Research Institute Genetics glossary (www.nhgri.nih.gov/DIR/VIP/Glossary/). The second source was the MeSH heading and subheadings contained within the MeSH tree heading "genetics" (i.e. MeSH tree number beginning with "G5"). All concepts from these two sources were normalized and mapped to UMLS using the KM concept indexer. We asked VUMS course directors and lecturers covering content in the first two years of the medical school to identify those lectures in their courses that contained a high number of genetics topics. We then took these "high genetics" lectures (n=23) and randomly selected ten for our study.

	Mean Total Unique Concepts	Mean Total Concept Matches	Mean% NIH Concept Matches	Mean% MeSH Concept Matches
High Genetics (n=10)	361.7± 163.2	721.7± 333.3	14.0± 6.9	6.0± 5.0
No Genetics (n=10)	416.2± 183.0	864.0± 456.6	2.7± 1.7	0.4± 0.5

Table 1. Genetics Concept matches from NIH and MeSH sources

We randomly selected ten lectures from the remaining corpus of digital lecture documents (n=650), classifying them as containing "no (minimal) genetics". Next, we submitted all twenty ASCII text lecture notes to the KM concept indexer. We calculated the total "genetics" concepts matched for both NIH and MeSH sources as well as the total unique concepts (i.e., total CUI's) and total concept matches for each document. After using KM to concept-index the entire preclinical curriculum, we identified all documents containing the selected MeSH genetics terms.

RESULTS

Results from concept indexing of the twenty lectures are shown in Table 1. The range of percent NIH concept matches in "high genetics" documents was 7.1% to 25.4% while in the "no genetics" documents it was 0.22% to 5.1%. The range of percent MeSH concept matches was 0% to 14.0% in "high" and 0.0% to 1.6% in "no genetics" documents. Percentage of document concepts indexed as MeSH concepts for the twenty study lectures ("high" and "no genetics") and for the entire VUMC preclinical curriculum (M1 and M2) are displayed in Figure 1.

DISCUSSION

In the small number of documents and concepts of this study there was a clear distinction between NIH match percentages for "high genetics" versus "no genetics" documents and only a slight overlap for MeSH match percentages for the two categories. While these matched concepts may not represent the only genetics concepts in the documents, they appear to be adequate "markers" for genetics content. Using a 1% cutoff for MeSH concept percentage, which represents the point of least overlap between "high" and "no genetics" documents, 93 documents representing 13.5% of the preclinical curriculum would be identified as "high genetics." We will test the validity of this categorization in future studies. If it is shown to be valid, this MeSH categorization could be adapted for the automated identification of other broad or focal subject content in the medical school curriculum. Thus, we believe this study indicates KM's future potential for assisting lecturers and school administrators in planning and evaluation of curricula.

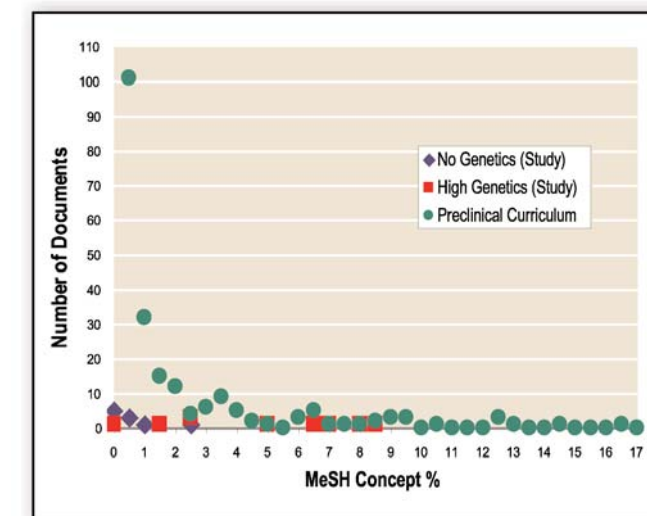


Figure 1. MeSH concept percentage for study documents and entire VUMC preclinical curriculum.

