Using MetaMap Program to Map Medical Curriculum to MCC and One45 Controlled Vocabulary

by

Nouf Bashwih
B00540138
nouf@cs.dal.ca

Performed at
Undergraduate Medical Education,
Clinical Research Centre
Room C-125, 5849 University Avenue
Dalhousie University
PO BOX 15000
Halifax, NS B3H 4R2

In partial fulfillment of the requirements of the Master of Health Informatics Program,
Dalhousie University

Report of Internship for the period May 9 – August 5, 2011

Date Submitted: August 29, 2011

Table of Contents

Acknowledgment 3
Executive Summary 4
1. Introduction 5
2. Undergraduate Medical Education 5
3. Project Description 6
3.1 Project Objective 6
3.2 MetaMap 7
3.3 Unified Medical Language System (UMLS) 7
3.4 Controlled Vocabulary 8
4. Work Performed 8
4.1 Determining the tasks and deadline 8
4.2 Cleaning the files 8
4.3 Renaming the files 8
4.4 Running the files in MetaMap program 8
4.5 Calculating the frequency 9
4.6 Mapping to MCC and One45 9
4.7 Creating the project package 9
5. How This Project Relates to Health Informatics 9
6. Critical Analysis of the Problem 10
7. Solution & Recommendations 12
8. Conclusion 12
References 14
APPENDIX A 15
APPENDIX B 20

Acknowledgment

This report has been written by me and has not received any previous academic credit at this or any other institution.

I wish to acknowledge Dr. Grace Paterson for her help and support during the undertaking of this project. Kevin French supervised the project to its completion, and the UGME staff provided assistance throughout the project—thanks to you all.

Nouf Bashwih

Executive Summary

This project was undertaken in order to assist Dalhousie University's UGME department in tracking the frequency at which particular concepts appear within its curriculum. Data was generated for the sake of analysis regarding the curriculum's strengths and weaknesses in terms of which concepts are being taught or discussed too frequently, and which ones are not being introduced frequently enough. This project mapped the lectures, cases and labs of Med 1 to UMLS using the MetaMap program. The materials were then mapped to MCC concepts and One45 controlled vocabulary. Through this process, the author measured the frequency with which specific medical concepts appeared in Dalhousie University's UGME curriculum. The main difficulty encountered was that not all of the concepts that appeared in the curriculum were mapped by the vocabularies employed. The author recommends using the MeSH vocabulary, in addition to MCC, in order to map the curriculum using a wider variety of concepts/terms. Mapping the curriculum with this broader vocabulary would result in improved precision. The author further recommends that Dalhousie University's UGME department consider developing its own controlled vocabulary in order to map the concepts/terms in the UGME curriculum with even greater precision. Mapping the curriculum with greater precision would generate more reliable data for curriculum analysis.

1. Introduction

The medical sector develops very quickly; therefore, heath care providers should have a solid background in order to be able to stay up to date with the latest developments. Medical school curriculum should be up to date in order for it to graduate well-prepared medical students who are able to effectively face real-life situations. Undergraduate Medical Education (UGME) departments have the authority and the responsibility to develop appropriate curriculum. The UGME aims to prepare students to be able to integrate their knowledge, skills and aptitudes (Undergraduate Medical Education, 2011).

Furthermore, UGME departments should be able to track their curriculum in order to know its strengths and weaknesses so they can focus more on the less-often covered areas and prevent the same materials from being covered in more than one class. This project has been developed in an effort to allow UGME departments to effectively track their curriculum. In the project, the lectures, cases and labs of Med 1 were mapped to the Unified Medical Language System (UMLS) using the MetaMap program, and were then mapped to Medical Council of Canada concepts (MCC) and One45 controlled vocabulary. This process was performed in order to measure the frequency with which specific medical concepts appear in the undergraduate curriculum.

Curriculum mapping facilitates the recognition of gaps and redundancies within curriculum. It also helps to elucidate where, when, and what is being taught (Willett, 2008). It is helpful for analyzing curriculum from multiple perspectives, and specific portions of curriculum, such as learning resource and timetables (Plaza, Draugalis, Slack, Shrepnek, & Sauer, 2007). Curriculum mapping has two main functions. The first function is to make the curriculum clearer to stakeholders, such as lecturers, students, curriculum planners, and managers. The second one is to reveal the links between curriculum components (Harden, 2001). Electronic curriculum mapping provides a comprehensive overview of medical curriculum, helps in curriculum searching and evaluation, and improves quality assurance processes. A study was conducted by Willett to research the current status of curriculum mapping in Canada and the UK, and it was found that the most used descriptors among the medical school in mapping are freeform keywords, controlled vocabularies, and MeSH (2008). In keeping with these findings, using electronic curriculum mapping and controlled vocabularies will help the UGME to achieve their goal of tracking their curriculum effectively.

2. Undergraduate Medical Education

Undergraduate Medical Education is under the Faculty of Medicine in Dalhousie University. It is responsible for developing and improving the medical curriculum. It aims to prepare the undergraduate students for their future degrees and careers through informative and challenging curriculum, aiming to foster creative and caring physicians. They enroll 109 new students yearly to prepare them to face real-life situations through helping them to develop aptitudes in four principle roles: professional, community contributor, lifelong learner and skilled clinician (Undergraduate Medical Education, 2011).

The four-year undergraduate program is divided into pre-clerkship years, which are years one and two, and clinical years, which are years three and four. The medical curriculum is designed in sequential units, which is different from a regular curriculum format, which is divided by terms. The Med 1 curriculum units are the Foundation of medicine, Host Defense, Metabolism and Homeostasis, Human Development, Professional Competencies, Clinical Skills I, Electives, and Rural week (Curriculum - MED 1, 2011).

The Foundation unit includes a foundation in biomedical, epidemiological, social, and human sciences. It is designed to help the students to complete the system-based units of the curriculum. It has two major parts (review of cell and molecular biology) and an introduction to evidence-based practices (Curriculum - MED 1, 2011).

The Host Defense unit includes the topics hematology, immunology, infection and inflammation. It helps students to develop knowledge about the blood components that are fighting the pathogens, the influence of pathogens nationally and internationally, the foundation of infections and infectious diseases, the immune system, and the role of blood cells and blood diseases (Curriculum - MED 1, 2011).

The third unit is Metabolism and Homeostasis, which includes two components: oral medicine and nutrition. The following unit is Human Development unit helps students build a solid foundation for Med II, Clerkship, and Residency in human reproduction, encompassing sexuality, the genitourinary system, genetics, embryology, labor and birth. Through clinical cases, students will illuminate and strengthen the achievement of basic concepts in anatomy and physiology. This unit also helps students to link these concepts with professionalism, patient-centeredness, and community responsibility (Curriculum - MED 1, 2011).

The Professional Competencies unit is a longitudinal two-year course taught in weekly tutorials. It helps students to combine biomedical and clinical learning within the context of patient care. It involves public health and infectious disease management in the community, end-of-life decision-making and other ethical challenges, patient safety and other system and quality improvement approaches, social accountability and global health, physician wellness and career paths, and the Health Mentors program (Curriculum - MED 1, 2011).

3. Project Description

The author has created a database of metadata for the UGME curriculum for Med 1. Medical concepts were identified in text documents using MetaMap. The MetaMap program maps text to UMLS concept IDs. After that, the author mapped the UMLS concepts to MCC objectives and One45 controlled vocabulary. The author used a space in the UGME office to perform her work, under the supervision of Kevin French, who is the document control technician.

3.1 Project Objective

As UGME aims to improve the medical curriculum, they are looking to overcome any current issues that they have in terms of the materials being taught. They want to be able to track the content of the curriculum, including lectures, labs and cases. As a result of this, they would be

informed about where each topic is taught, what is neglected within the curriculum, and where there are overlaps in the material being taught. Therefore, this project has been developed in order to assist UGME in achieving their goals.

3.2 MetaMap

In this project, the author used the MetaMap program to map the curriculum concepts to UMLS. Dr. Alan Aronson developed a program called Metamap, which discovers Metathesaurus concepts referred to in any particular text. Metamap uses information based on representative, natural language and figures linguistic techniques. Metamap is considered fundamental in developing both semiautomatic and fully automatic indexing of biomedical literature at the National Library of Medicine (NLM) (About MetaMap, 2011).

MetaMap works by dividing the text into phrases. It then maps each phrase, with mapping options returned from each phrase ranked according to their strength. MetaMap is used by many researchers in text mining, information retrieval, and specific concepts extraction such as anatomical terms (Pratt & Yetisgen-Yildiz, 2003).

3.3 Unified Medical Language System (UMLS)

Unified Medical Language System (UMLS) is a set of files and software that contains many health and biomedical vocabularies and standards. It aims to facilitate interoperability between computer systems. UMLS has many uses, such as enhancing or developing applications, including electronic health records, classification tools, dictionaries and language translators (Unified Medical Language System, 2011).

UMLS links different computer systems' health information, medical terms, drug names and billing codes together. It enables this through connecting codes and terms between doctors, pharmacies, and insurance companies, and by coordinating patient care between hospital departments. Furthermore, UMLS has functions that include search engine repossession, data withdrawal, public health information reporting, and terminology research (Unified Medical Language System, 2011).

UMLS has three tools that are knowledge sources. The first tool is metathesaurus, which has many terms and codes such as CPT, ICD-10-CM, LOINC, and RxNorm. It also has a semantic network tool with semantic types and relationships. The third tool is a specialist lexicon and lexical tools, which comprises natural language processing tools (Unified Medical Language System, 2011) (understanding).

UMLS is very comprehensive since it includes many controlled vocabularies such as MeSH, SNOMED-CT, and Gene Ontology (Ruau, Mbagwu, Dudley, Krishnan, & Butte, 2011). One important criteria for detecting any neglect or redundancy in curriculum design is the ability to retrieve concepts with satisfactory precision and sensitivity (Kanter, 1993). Therefore, the author chose to map to UMLS instead of MeSH or SNOMED-CT, because UMLS is more comprehensive and more sensitive since it contains MeSH and SNOMED-CT concepts, and thus able to capture more concepts.

3.4 Controlled Vocabulary

A controlled vocabulary consists of a controlled list of terms used by users to map concepts. It can be useful when searching for specific topics, themes, or concepts within curriculum (Willett, 2008). One example of a controlled vocabulary is One45 controlled vocabulary developed at the University of Alberta (French, 2011). The UGME wanted to map their curriculum to One45 controlled vocabulary, in order to help with the tracking of their curriculum. The One45 controlled vocabulary has five distinct content areas. The first one is the area of study and skills, such as behavioural and social sciences. The second one is body systems like the cardiovascular system. The third content area includes fields of study, such as anatomy. The fourth is health conditions, such as gastroenterology. The fifth and final content area is symptoms, such as fever.

4. Work Performed

4.1 Determining the tasks and deadline

The author met with her supervisor in order to determine the tasks to be performed and the deadlines for the completion of each task. During the meetings, the author and her supervisor agreed upon the deliverables (see Appendix A).

4.2 Cleaning the files

The author cleaned 300 files, which included the course objectives, lectures and cases before running them through MetaMap. This cleaning process involved removing from all documents numbers and demarcations such as dates, values and punctuation, and characters such as semi-colons and brackets. Also, the author removed all extra spacing and names of places or people. After being cleaned, the files were ready for MetaMapping.

4.3 Renaming the files

The author changed the name of lectures to numbers to make it easier to organize and to prevent any error in writing the name of lectures when running them in the MetaMap program. For example, the number 10 is related to 20101004_MED1_FUN_DW_DESIGN_AND DISCOVER_OF CURATIVE_DRUGS_LEC. It appears as 10_ 20101004_MED1_FUN_DW_DESIGN_AND DISCOVER_OF CURATIVE_DRUGS_LEC.

4.4 Running the files in MetaMap program

The author ran the files through the Metamap program to map them onto UMLS by writing the following command:
nouf@hector:~ /internship metamap10 –I example.txt example_umls.txt

4.5 Calculating the frequency

The author calculated the frequency of how often concepts appeared in order to measure which concepts were mentioned more often and which were neglected. The author used the Medical Curriculum Mapping Program to achieve this. After that, the author sorted the concepts in descending order.

4.6 Mapping to MCC and One45

The next step was to map the top twenty concepts for each session in the Med 1curriculum to MCC and One45 controlled vocabulary manually. This mapping process revealed which concepts are repeatedly mentioned and which concepts are rarely mentioned.

4.7 Creating the project package

The final step was to create the project package, which contains the combined Excel files, the separate files, and the name directory. The combined file includes all mapped files in one Excel file, organized by numbers. The separate files are organized in separate documents according to their courses. The name directory is created by Excel and contains all of the lecture names and numbers.

5. How This Project Relates to Health Informatics

Health Informatics involves applying information technology in order to create and make use of health-related data, information and knowledge. Health Informatics as a discipline helps ensure that patients receive safer, more efficient and more effective healthcare. Professionals in the field of Health Informatics draw from fields such as computer science, information management, cognitive science, communications, epidemiology, management sciences and health sciences in order to develop and deploy information and systems solutions. One of the major roles of health informatics is to develop classification systems using standardized terminology and coding (About Health Informatics, n.d.).

This particular Health Informatics project uses the computer program MetaMap in order to map curriculum texts to UMLS concepts. The author then used the Medical Curriculum Mapping Program to analyze how frequently certain medical concepts appeared in the undergraduate curriculum. Thus, computer science and information management were used in combination with the health sciences curriculum to develop information that could offer solutions to curriculum issues.

Through her university education at Dalhousie University, the author gained helpful information regarding Health Informatics, which informed this project. In particular, she learned a great deal regarding MetaMap and UMLS through her participation in a Health Informatics Flow and Standards course taken with Dr. Grace Paterson in the Winter semester of 2011. In completing one of the assignments for that course, the author used the MetaMap program to map a case to UMLS, MeSH, and SNOMED CT concepts. Through this experience, the author learned about the precision and sensitivity of these concepts, and about how to use the MetaMap program effectively. This educational and experiential background in Health Informatics was applied by the author over the course of completing the present project.

6. Critical Analysis of the Problem

As mentioned previously, the UGME wanted to solve their problem concerning the difficulty of tracking their curriculum, through mapping it to MCC and One45 controlled vocabulary. However, some of the lectures pertaining to topics such as statistics, patient safety and physician-patient communication did not map to MCC or One45 controlled vocabularies because these vocabularies are too narrow. The MCC concepts are mostly symptoms of illness or disease, such as pain, fever or nausea. The One45 controlled vocabulary was developed by the University of Alberta based on their curriculum; while extensive, it does not cover all areas included in the Med I curriculum at Dalhousie University (French, 2011). These two vocabularies allowed the author to map a great number of concepts that appear within the specified undergraduate curriculum, but as the author was mapping the concepts, she recognized that many concepts were not included in these vocabularies. These unmapped concepts included common diseases such as Alzheimer's, and congenital anomalies. It was not within the scope of this project to work outside of the MCC and One45 vocabularies, so many concepts in the curriculum were left unmapped.

Upon analyzing the mapped data, the author found that the mapping process revealed how terms were repeated in numerous lectures but without the material being repeated. This was the case when the same subject was used repeatedly in examples. Cancer and Human Immunodeficiency Virus (HIV) were the diseases that seemed to be mentioned most often in the lectures when an example was being made. Thus, the concept of cancer and HIV were mapped across many different lectures, but these lectures only mentioned cancer and HIV in their examples—they were not repeating the same lessons.

MCC concepts, as previously mentioned, are narrow concepts. As can be seen in Appendix B, there are many concepts that fit in multiple maps (MM) because they are broad concepts. For example, the concept Cancer is covered in MMC in multiple areas, including are Prostatic Cancer and Prevention of Lung Cancer. This resulted in the creation of one-to-many mappings for broad concepts. On the other hand, there are narrow concepts such as Sudden Infect Death Syndrome or Diabetes Mellitus, which map one-to-one because they are narrow concepts.

The author conducted a search to find similar or different approaches to concept mapping presented in literature about medical curriculum, and the findings of this research. Denny, Smithers, Miller, and Spickard (2003) developed a study to evaluate the functionality of a prototype version of KnowledgeMap (KM) program for identifying medical concepts, as compared to MetaMap when

set to default options with quick composites, and using UMLS strict model to ignore stop phrases. The KM uses the SPECIALIST lexicon and Metathesaurus derived from the UMLS, heuristic language processing techniques, and an empirical scoring algorithm. The authors took the lectures of the first two years of Vanderbilt Medical School and recognized the important concepts manually, which they called a "gold standard." They found that MetaMap identified 78% of the concepts while KM identified 82%, with precisions being 85% and 89% respectively. The reason that they provided for matching failures was a lack of target concepts from UMLS Metathesaurus (Denny, Smithers, Miller, & Spickard, 2003). The similarity between this study and the present one is that they both used MetaMap to map the medical curriculum. However, the difference is that Denny, Smithers, Miller, and Spickard wanted to measure the ability of KM and compare it with MetaMap, and they found that KM is more precise and sensitive than MetaMap.

There is a study that was carried out by Denny, Bastarache, Sastre, and Spickard III (2009) to track the medical curriculum using KM to find UMLS concepts. The researchers used clinical notes from three finishing fourth year students to analyze the ability of an automated algorithm to identify 10 core clinical problems. It was found that recall and precision for UMLS concept identification was approximately 0.91 and 0.92. This study was limited to certain circumstances, and it may not be applicable for clerkships where clinical notes are not the main measure of performance and learning, such as surgical clerkships (2009). This study is related to the present study insofar as they used UMLS to track the medical curriculum, while the difference is the type of program used to map. Denny, Bastarache, Sastre, and Spickard III (2009) used KnowledgeMap while the author used MetaMap.

Kanter (1993) followed an automated method to represent concepts from medical school lectures using UMLS resources. They chose two different lectures from a first year medical school curriculum. Different terms were chosen from each lecture outline, depending on their relevance to the lecture topic, which were also MeSH main headings presented in UMLS metathesaurus. Information was then collected for each term selected, such as the total number of indexed articles that had a particular concept as a main heading in MEDLINE during a specific period of time, and the number of co-occurring terms. UMLS resources were found to help identify the relationship between information in medical curriculum and biomedicine. Moreover, this automated system can assist faculty in planning lectures and courses, or it can help support decisions regarding making alterations to the curriculum. This study differs from this paper in that Kanter's method was to use UMLS resources in mapping the medical curriculum, while in this paper the author used UMLS concepts.

This project was mainly devoted to the generation of data regarding the frequency at which concepts occur in the curriculum for Dalhousie University's undergraduate medical program. It is beyond the scope of this project to analyze or base a critique on the frequency at which different concepts appear in the curriculum. It will be the responsibility of the Dalhousie University UGME administrators and faculty to analyze the data and decide if certain concepts are being taught and discussed too frequently, not frequently enough, or at an appropriate frequency.

7. Solution & Recommendations

The main problem that this project faced is the limited vocabulary offered by MCC and One45 for mapping purposes. As noted above, these two datasets did not include all of the concepts that are included in Dalhousie University's UGME curriculum. Thus, many concepts were left unmapped. The solution that the author proposes for this problem is that the Medical Subject Heading (MeSH) dataset also be used to map the concepts that appear in the curriculum.

MeSH is a controlled vocabulary thesaurus of the NLM. It allows searching at a range of specificity by having a hierarchical structure of terms, naming descriptors which are organized in an alphabetical and hierarchical structure. It has a twelve-level hierarchy, 26,142 descriptors, and 177,000 entry terms that help to find the most suitable MeSH heading. Moreover, it has more than 199,000 headings that are referred to as Supplementary Concept Records in a different thesaurus (Medical Subject Headings [MeSH®] Fact Sheet, 2011).

The MeSH has been used in many areas. For instance, it is used by NLM to index biomedical journals for MEDLINE and PubMED and for the cataloging of books, documents, and audiovisuals in NLM-produced database. Moreover, it helps to find desired topics by using MeSH in search queries (Medical Subject Headings [MeSH®] Fact Sheet , 2011).

MeSH is considered a good solution, since it includes many concepts that MCC and One45 do not. Moreover, it is accessible through Dalhousie University's One45 website, and the author feels that MeSH will likely map the concepts more precisely than MCC and One45 because of its broader vocabulary. The author recommends that this additional mapping be conducted before rigorous analysis of the concepts in the UGME curriculum is undertaken. Moreover, since MCC objectives should be used to determine the extent to which MCC objectives are included in UGME curriculum, MCC should be used in addition to MeSH.

The author further recommends that, in the future, Dalhousie University's UGME could develop its own controlled vocabulary of terms/concepts that it feels should be included in its curriculum. Creating such a unique vocabulary would allow for even more precise mapping of the curriculum, and thus more accurate data for analysis. Despite the previously mentioned advantages of using controlled vocabularies, many challenges face anyone aiming to build a new one. The most cited problems are the requirements of ample time and human resources (Willett, 2008). Moreover, controlled vocabularies require regular maintenance to compensate for any new updates, and this process is costly and time consuming (Kanter, 1993).

8. Conclusion

This project was undertaken in order to assist Dalhousie University's UGME department in tracking its curriculum. Data was generated for the sake of analysis regarding the curriculum's strengths and weaknesses in terms of what concepts are being taught or discussed too frequently and which ones are not being introduced frequently enough. This project mapped the lectures, cases and labs of Med 1 to UMLS using the MetaMap program. The materials were then mapped to MCC concepts and One45 controlled vocabulary. Through this process, the author measured the

frequency with which specific medical concepts appeared in Dalhousie University's UGME curriculum. Not all of the concepts that appeared in the curriculum were mapped by the vocabularies employed, and this report includes recommendations concerning how the curriculum could be mapped more precisely.

References

About Health Informatics. (n.d.). Retrieved August 9, 2011, from Canada's Health Informatics Association: http://www.coachorg.com/health_informatics/about_health_informatics.htm

About MetaMap. (2011, August). Retrieved August 6, 2011, from MetaMap Portal: http://mmtx.nlm.nih.gov/

Curriculum - MED 1. (2011, August 12). Retrieved August 1, 2011, from Dalousie University Faculty of Medicine: http://undergraduate.medicine.dal.ca/curriculum/med1.htm

Denny, J. C., Bastarache, L., Sastre, E. A., & Spickard III, A. (2009). Tracking medical students' clinical experiences using natural language processing. *Journal of Biomedical Informatics , 42* (5), 781-9.

Denny, J. C., Smithers, J. D., Miller, R. A., & Spickard, A. (2003). "Understanding" medical school curriculum content using KnowledgeMap. *Journal of the American Medical Informatics Association , 10* (4), 351-62.

French, K. (2011). Document control technician, Undergraduate Medical Education, Dalhousie University, Halifax, Nova Scotia, Canada.

Harden, R. (2001). AMEE Guide No. 21: Curriculum mapping: a tool for transparent and authentic teaching and learning. *Medical Teacher , 23* (2), 123-137.

Kanter, S. L. (1993). *Using the UMLS to represent medical curriculum content* . Pittsburgh: Proc Annu Symp Comput Appl Med Care.

Medical Subject Headings (MeSH®) Fact Sheet . (2011, January 6). Retrieved August 13, 2011, from U.S. National Library of Medicine:http://www.nlm.nih.gov/pubs/factsheets/mesh.html

Pratt, W., & Yetisgen-Yildiz, M. (2003). *A study of biomedical concept identification: MetaMap vs. people.* Retrieved August 8, 2011, from http://www.ncbi.nlm.nih.gov/pubmed/1472822

Ruau, D., Mbagwu, M., Dudley, J. T., Krishnan, V., & Butte, A. J. (2011, March 21). Comparison of automated and human assignment of MeSH terms on publicly-available molecular datasets. *Journal of Biomedical Informatics*.

Undergraduate Medical Education. (2011, July 4). Retrieved July 30, 2011, from Dalhousie University: http://undergraduate.medicine.dal.ca/curriculum/index.htm

Unified Medical Language System. (2011, April 15). Retrieved August 8, 2011, from U.S. National Library of Medicine: http://www.nlm.nih.gov/research/umls/quickstart.html

Plaza, C. M., Draugalis, J. R., Slack, M. K., Shrepnek, G. H., & Sauer, K. A. (2007). Curriculum mapping in program assessment and evaluation. *American Journal of Pharmaceutical Education , 71* (2).

Willett, T. G. (2008). Current status of curriculum mapping in Canada and the UK. *Med Educ.* , 786-93.

APPENDIX A

Proposed Work Plan

TASK	WORK HOURS (MIN-MAX)		TARGET DATE	DELIVERABLE
1. Clean the Foundation files for metamapping.	32	36	May 26, 2011	Cleaned files.
2. Clean the Host Defence files for metamapping.	90	111	June 13, 2011	Cleaned files.
3. Run both documents in MetaMap program.	10	12	June 14, 2011	Mapped files to UMLS.
4. Clean the Metabolism & Homeostasis files for metamapping.	50	55	June 19, 2011	Cleaned files.
5. Clean the Human Development files for metamapping.	35	40	June 24, 2011	Cleaned files.
6. Clean the ProComp files for metamapping.	3	5	June 27, 2011	Cleaned files.
7. Clean the Rural week objective for metamapping.	1	2	June 27, 2011	Cleaned files.
8. Clean the Elective objective for metamapping.	1	2	June 27, 2011	Cleaned files.

TASK	WORK HOURS (MIN-MAX)		TARGET DATE	DELIVERABLE
9. Clean the Clinical Skill objective for metamapping.	1	2	June 27, 2011	Cleaned files.
10. Calculate the frequency for each file of Foundation and organize them descending.	3	4	June 27, 2011	Files ready to map to One45-controlled vocabulary and MCC.
11. Calculate the frequency for each file of Host Defence and organize them in descending order.	4	5	June 28, 2011	Files ready to map to One45-controlled vocabulary and MCC.
12. Calculate the frequency for each file of Metabolism & Homeostasis and organize them in descending order.	6	7	June 29, 2011	Files ready to map to One45-controlled vocabulary and MCC.
13. Calculate the frequency for each file of Human Development and organize them in descending order.	5	7	June 30, 2011	Files ready to map to One45-controlled vocabulary and MCC.

TASK	WORK HOURS (MIN-MAX)		TARGET DATE	DELIVERABLE
14. Calculate the frequency for each file of ProComp and organize them in descending order.	1	2	June 30, 2011	Files ready to map to One45-controlled vocabulary and MCC.
15. Calculate the frequency for each file of Rural week, Elective and Clinical Skill, organizing them in descending order.	1/2	1	July 1, 2011	Files ready to map to One45-controlled vocabulary and MCC.
16. Map the Foundation files to One45 and MCC.	50	56	July 8, 2011	Mapped files.
17. Map the Host Defence files to One45 and MCC.	70	75	July 15, 2011	Mapped files.
18. Map the Metabolism & Homeostasis files to One45 and MCC.	90	96	July 22, 2011	Mapped files.

TASK	WORK HOURS (MIN-MAX)		TARGET DATE	DELIVERABLE
19. Map the Human Development files to One45 and MCC.	60	63	August 2, 2011	Mapped files.
20. Map the ProComp files to One45 and MCC.	2	4	August 4, 2011	Mapped files.
21. Map the Rural, Elective and Clinical Skills files to One45 and MCC.	2	3	August 4, 2011	Mapped files.
22. Create the project package which contains combined files, separate files, and name directory.	6	8	August 5, 2011	Project package ready for analysis and upload.
Total	522	596		

APPENDIX B
The Top Concepts in Each Unit

Concepts	Concept CUIs	***Foundation Unit***	Concept Category	Concept Descrition	MCC	One45
Energy	C0542479	69	Natural Phenomenon or Process	Energy, Physics		
Injury	C0175677	57	Injury or Poisoning		Trauma/Accidents	
Haemoglobin	C0019046	57	Amino Acid, Peptide, or Protein,Biologically Active Substance	Hemoglobin	Hemoglobin Serum, Abnormal, Elevated Hemoglobin.	Hemic System
vaccine	C0042210	53	Immunologic Factor,Pharmacologic Substance	Vaccines	Infant and Child Immunization	Pharmacology and Pharmacotherapeutics
???lead	C0023175	44	Element, Ion, or Isotope,Hazardous or Poisonous Substance	Lead	Poisoning	
Feedback	C0015744	43	Phenomenon or Process			
Genome	C0017428	41	Gene or Genome		Genetic Concerns	Medical Genetics
Hazards	C0598697	34	Qualitative Concept	Hazard		
Cyclosporin A	C0010592	33	Amino Acid, Peptide, or Protein,Pharmacologic Substance	Cyclosporine		Pharmacology and Pharmacotherapeutics
Population Programs	C0032678	32	Health Care Activity		Population Health	
Reflux	C0232483	30	Pathologic Function			Gastroesophageal Reflux Disease
Cells	C0007634	30	Cell			Histology
Genetic mutation	C0026882	28	Genetic Function	Mutation	Genetic Concerns	Medical Genetics

Drugs	C0013227	26	Pharmacologic Substance	Pharmaceutical Preparations		Pharmacology
Evidence	C0332120	24	Functional Concept	Evidence of		
Sudden infant death syndrome	C0038644	23	Disease or Syndrome		Sudden Infant Death Syndrome (SIDS)	
Diabetes	C0011849	21	Disease or Syndrome	Diabetes Mellitus	Diabetes Mellitus	Diabetes Mellitus
Cancer	C0006826	20	Neoplastic Process	Malignant Neoplasms	Prostatic Cancer, Prevention of Lung Cancer.	Oncology
Proteins	C0033684	20	Amino Acid, Peptide, or Protein,Biologically Active Substance		Proteinuria, Pregnancy Associated Hypertension.	Biochemistry
Validity	C2349101	20	Quantitative Concept	Validity (characteristic)		
*^patient	C0030705	18	Patient or Disabled Group	Patients		
Lymphatic Invasion	C1708790	18	Finding			
Signal Transduction	C0037083	17	Cell Function			Cell Biology
Enzymes	C0014442	16	Enzyme,Organic Chemical		Jaundice, Pregnancy Associated Hypertension.	Biochemistry
Aspirin	C0004057	16	Organic Chemical,Pharmacologic Substance			Pharmacology and Pharmacotherapeutics
Evidence	C0332120	15	Functional Concept	Evidence of		
Association	C0439849	14	Qualitative Concept	Relationships		

Laboratory	C0022877	14	Manufactured Object,Organization			
Membranes	C0025255	14	Tissue	Tissue membrane		Histology
Bias	C0242568	14	Idea or Concept	Biases		
Alzheimer's dementia	C0002395	11	Disease or Syndrome	Alzheimer's Disease	Dementia	Dementia
Prevalence	C0033105	11	Quantitative Concept	Statistical Prevalence		
Mutations	C0026882	10	Genetic Function	Mutation	Genetic Concerns	Medical Genetics
Vessels	C0005847	10	Body Part, Organ, or Organ Component	Blood Vessels	Pulse Abnormalities	Cardiovascular system
Gene Expression	C0017262	9	Genetic Function		Genetic Concerns	Medical Genetics
Evidence-Based Medicine	C0376537	8	Biomedical Occupation or Discipline			
Patient Safety	C2362562	7	Intellectual Product	Patient Safety Topics		
Variable	C1705098	7	Quantitative Concept	Variable (statistical)		
Confidence Intervals	C0009667	6	Quantitative Concept			
Infections	C0021311	4	Disease or Syndrome	Infection	Abdominal distension, Abdominal pain, blood in urine.	
Body	C1268086	4	Anatomical Structure	Body structure		
Attributable Risk	C0814766	3	Qualitative Concept			
Risk Ratio	C0028873	2	Quantitative Concept	Odds Ratio		

CPOE	C0973138	2	Medical Device	Computerized Physician Order Entry System		

Concepts	**Concept CUIs**	***Host Defence Unit***	**Concept Category**	**Concept Descrition**	**MCC**	**One45**
Malaria	C0024530	90	Disease or Syndrome		Bites, Animal/Insects	
Red Blood Cells	C0014792	76	Cell	Erythrocytes	Anemia, Blood From Gastrointestinal Tract.	Hemic System
Diabetes	C0011849	72	Disease or Syndrome	Diabetes Mellitus	Diabetes Mellitus	Diabetes Mellitus
Iron	C0302583	71	Biologically Active Substance,Element, Ion, or Isotope,Pharmacologic Substance		Anemia	Biochemistry
Haemostasis, NOS	C0019116	64	Organ or Tissue Function	Hemostatic function	Hematemesis	
VON WILLEBRAND DISEASE	C1336938	59	Gene or Genome	VWF gene		von Willebrand Disease
Influenza	C0021400	55	Disease or Syndrome		Fever in the Immune Compromised Host	Influenza
Antigen	C0003320	53	Immunologic Factor	Antigens		Immune System
HIV infection	C0019693	53	Disease or Syndrome	HIV Infections	Fever in the Immune Compromised Host	Human Immunodeficiency Virus
Shingles	C0019360	52	Disease or Syndrome	Herpes zoster disease	Fever in the Immune Compromised Host	Herpes
Asthma	C0004096	50	Disease or Syndrome		Respiratory Difficulty (Asthma)	Asthma
T cell	C0039194	47	Cell	T-Lymphocyte		Immune System
HIV	C0019682	46	Virus		Fever in the Immune	Human Immunodeficie

					Compromised Host	ncy Virus
Meningitis	C0025289	45	Disease or Syndrome		Headache	Meningitis
Antibody	C0003241	44	Amino Acid, Peptide, or Protein,Immunologic Factor	Antibodies		Immune System
Spleen	C0037993	44	Body Part, Organ, or Organ Component		Splenomegaly	
Cancer	C0006826	39	Neoplastic Process	Malignant Neoplasms	Prevention of Lung Cancer, Prostatic Cancer	Oncology
Infections, Urinary Tract	C0042029	39	Disease or Syndrome	Urinary tract infection		Urinary tract infection
Bleeding	C0019080	36	Finding	Hemorrhage	Vaginal Bleeding, Bleeding Tendency	Bleeding
Deep-Venous Thrombosis	C0149871	34	Disease or Syndrome	Deep Vein Thrombosis	Unilateral/ Local Edema	Deep Venous Thrombosis
Cancer	C0027651	33	Neoplastic Process	Neoplasm	Prevention of Lung Cancer, Prostatic Cancer	Oncology
Gene	C0017337	32	Gene or Genome	Genes	Genetic Concern	Medical Genetics
Organisms	C0029235	32	Organism	Organism		Medical Microbiology
Blood	C0005767	31	Tissue		Blood From Gastrointestinal Tract, Blood in urine, Blood in sputum.	Hemic System
Deep Vein Thrombosis	C0149871	31	Disease or Syndrome		Unilateral/ Local Edema	Deep Vein Thrombosis
Drug	C0013227	30	Pharmacologic Substance	Pharmaceutical Preparations		Pharmacology and Pharmacotherapeutics

Vaccines	C0042210	30	Immunologic Factor,Pharmacologic Substance		Infant and Child Immunization	Pharmacology and Pharmacotherapeutics
Parasites	C0030498	29	Eukaryote			Medical Micobiology
virus infection	C0042769	28	Disease or Syndrome	Virus Diseases	fever in the Immune Compromised Host	Medical Microbiology
SOFT TISSUE INFECTION	C0149778	28	Disease or Syndrome	Soft Tissue Infections	Fever in the Immune Compromised Host	
Folate	C0178638	26	Organic Chemical,Pharmacologic Substance,Vitamin			Biochemistry
Specimen	C2347026	26	Conceptual Entity	Biospecimen		
LEUKAEMIA	C0023418	26	Neoplastic Process	leukemia	Anemia, Periarticular Pain/Soft Tissue Rheumatic Disorders, Hepatomegaly, Splenomegaly	leukemia
Lymph node	C0024204	25	Body Part, Organ, or Organ Component	lymph nodes	Lymphadenopathy	Immune System
Bacteria	C0004611	25	Bacterium		Diarrhea, Dyspnea, Abdominal pain, Eye redness	Medical Microbiology
Pneumonia	C0032285	24	Disease or Syndrome		Blood in sputum, Cough, Cyanosis/Hypoxemia/Hypoxia	Pneumonia
immune deficiency	C0850497	22	Disease or Syndrome		Fever in the Immune Compromised Host	Immunology

Antibiotic	C0003232	21	Antibiotic	Antibiotics		Pharmacology and Pharmacotherapeutics
Lupus Erythematosus, Systemic	C0024141	21	Disease or Syndrome		Joint Pain, Alopecia, Pleural Effusion/Pleural Abnormalities.	Systemic Lupus Erythematosus
Gonorrhea	C0018081	21	Disease or Syndrome		Vaginal Discharge/Vulvar Itch/STD	Gonorrhea
Human Papillomavirus	C0021344	21	Virus		Vaginal Discharge/Vulvar Itch/STD	Human Papillomavirus
Rejection	C1882923	20	Phenomenon or Process	Rejection Due to Device-Patient Incompatibility		
Staphylococcus	C0038170	19	Bacterium	Genus staphylococcus		Medical Microbiology
Skin	C1123023	17	Body System			Integumentary System
Laboratory Tests	C0022885	17	Laboratory Procedure	Laboratory Procedures		
Fungi	C0016832	14	Fungus	fungus	Fever in the Immune Compromised Host	Medical Microbiology
Immunoglobulins	C0021027	14	Amino Acid, Peptide, or Protein,Immunologic Factor,Pharmacologic Substance			Immunology
Chemotherapy	C0013216	12	Therapeutic or Preventive Procedure	Pharmacotherapy		Pharmacotherapy
Neutrophile	C0027950	12	Cell	neutrophil		Immune System
Urticaria	C0042109	12	Disease or Syndrome		Urticaria	

Cyclooxygenase	C0389003	11	Amino Acid, Peptide, or Protein,Enzyme	cyclooxygenase 1		Biochemistry
Lymphatic	C0229889	11	Body Part, Organ, or Organ Component	Lymphatic vessel		Immune System
Lymphoma	C0024299	10	Neoplastic Process		Abdominal mass, splenomegaly	Lymphoma
Asthma	C0004096	9	Disease or Syndrome		Respiratory Difficulty (Asthma)	Asthma
Informed Consent	C0021430	9	Regulation or Law			Informed Consent
Blood Transfusion	C0005841	9	Therapeutic or Preventive Procedure			
Acute myeloid leukaemia	C0023467	9	Neoplastic Process	Leukemia, Myelocytic, Acute		Leukemia
Lymphoma	C0024299	8	Neoplastic Process		Abdominal mass, splenomegaly	Lymphoma
Strep	C0038395	7	Disease or Syndrome	Streptococcal Infections		Medical Microbiology
Allergy	C0020517	6	Pathologic Function	Hypersensitivity	Allergic Reaction	Allergy
Infection	C0021311	6	Disease or Syndrome		White Blood Cells, Abnormalities of	

Concepts	Concept CUIs	***Metabolism & Homeostasis Unit***	Concept Category	Concept Descrition	MCC	One45
diabetes	C0011860	107	Disease or Syndrome	Diabetes Mellitus, Non-Insulin-Dependent	Diabetes Mellitus	Diabetes Mellitus
Obesity	C0028754	104	Disease or Syndrome		Obesity	Obesity
Diabetes	C0011849	79	Disease or Syndrome	Diabetes Mellitus	Diabetes Mellitus	Diabetes Mellitus
Fractures	C0016658	70	Injury or Poisoning	Fracture	Fracture	Fracture
Glucose	C0017725	67	Biologically Active Substance,Carbohydrate,Pharmacologic Substance		Hypocalcemia , Hypercalcemia , Diabetes mellitus	Biochemistry
Iodine	C0021968	64	Element, Ion, or Isotope,Pharmacologic Substance		Thyroid Disease	Biochemistry
Elastase	C0699918	63	Amino Acid, Peptide, or Protein,Enzyme	Elastases		Biochemistry
Glucose	C0017725	55	Biologically Active Substance,Carbohydrate,Pharmacologic Substance		Hypocalcemia , Hypercalcemia , Diabetes mellitus	Biochemistry
Risk	C0035647	54	Qualitative Concept			
Cirrhosis	C0023890	51	Disease or Syndrome	Liver Cirrhosis	Jaundice, Liver Function Tests Abnormal, Serum	Cirrhosis

Oral Cancer	C0153381	49	Neoplastic Process	Malignant neoplasm of mouth		Upper Aerodigestive Tract Cancers
Hypercortisolism	C0010481	49	Disease or Syndrome	Cushing Syndrome	Adrenal Mass	Cushing Syndrome
DIARRHOEA	C0011991	49	Sign or Symptom	Diarrhea	Diarrhea	Diarrhea
Leucoplakia	C0023531	49	Neoplastic Process	Leukoplakia		Oncology
Gastresophageal reflux disease	C0017168	44	Disease or Syndrome	Gastroesophageal reflux disease		Gastroesophageal reflux disease
Graves	C0018213	44	Disease or Syndrome	Graves Disease	Thyroid Disease	Graves Disease
Teeth	C0040426	43	Body Part, Organ, or Organ Component	Tooth structure		Dentistry
Cholesterol	C0008377	42	Biologically Active Substance,Steroid		Lipids Abnormal, Serum	Biochemistry
pheochromocytoma	C1706920	42	Neoplastic Process	Benign pheochromocytoma of adrenal gland		Oncology
Obesity	C0028754	41	Disease or Syndrome		Obesity	Obesity
Hormones	C0019932	40	Hormone		Stature Abnormal, Menopause	Endocrine System
cholesterol synthesis	C0599474	40	Molecular Function	cholesterol biosynthetic process		
Pituitary	C0032005	37	Body Part, Organ, or Organ Component	Pituitary Gland		Endocrine System

Pancreas	C0030274	36	Body Part, Organ, or Organ Component		Diabetes mellitus, Abdominal mass	Gastrointestinal System
Gastric Cancer	C0699791	36	Neoplastic Process	Stomach Carcinoma	Abdominal Pain, Chronic	Upper Aerodigestive Tract Cancers
COPD	C0024117	35	Disease or Syndrome	Chronic Obstructive Airway Disease	Chronic dyspnea	Chronic Obstructive Pulmonary Disease
Liver	C0023884	33	Body Part, Organ, or Organ Component		Hepatomegaly, Liver Function Tests Abnormal, Serum	Gastrointestinal System
Pain	C0030193	33	Sign or Symptom		Pain	
Primary sclerosing cholangitis	C0566602	33	Disease or Syndrome		Acute abdominal pain, Jaundice	Cholangitis
Insulin	C0021641	32	Amino Acid, Peptide, or Protein,Hormone,Pharmacologic Substance		Diabetes mellitus, hypoglycemia	Endocrine System
esophageal motility	C0920509	31	Organ or Tissue Function	Esophagus motility		Gastrointestinal System
Gastrointestinal Tract	C0017189	31	Body System	Gastrointestinal tract structure		Gastrointestinal System
Cortisol	C0020268	30	Hormone,Pharmacologic Substance,Steroid	Hydrocortisone	Hypertension, Osteoporosis	Endocrine System
Sex Characteristics	C0036866	30	Organism Attribute			

Endoscopy	C0014245	29	Diagnostic Procedure	Endoscopy (procedure)		Diagnostic Imaging
colonoscopy	C0009378	29	Diagnostic Procedure			Diagnostic Imaging
Fluoride	C0016327	29	Inorganic Chemical,Pharmacologic Substance	Fluorides		
Carbohydrate	C0007004	28	Carbohydrate	Carbohydrates		Biochemistry
Early childhood caries	C1290630	28	Disease or Syndrome	Caries of infancy	Mouth problem	Dentistry
Gonadal	C0018067	28	Body Part, Organ, or Organ Component	Gonadal structure		Reproductive System
Ionized Calcium	C0596235	27	Element, Ion, or Isotope	Calcium ion	Hypercalcemia	Biochemistry
Abdominal	C0000726	26	Body Location or Region	Abdomen		
Liver diseases	C0023895	25	Disease or Syndrome		Hepatomegaly, Liver Function Tests Abnormal, Serum	Gastrointerology
Colon cancer	C0699790	24	Neoplastic Process	Colon Carcinoma	Chronic diarrhea	Colon cancer
Fracture	C0016658	23	Injury or Poisoning		Fracture	Fracture
Pituitary	C0032005	23	Body Part, Organ, or Organ Component	Pituitary Gland		Endocrine System
Celiac Disease	C0007570	22	Disease or Syndrome			Celiac Disease
Ascites	C0003962	22	Finding			
Thyroid cancer	C0007115	21	Neoplastic Process	Malignant neoplasm of thyroid	Neck Mass/Goiter/ Thyroid Disease	Thyroid Cancer

Crohn Disease	C0010346	20	Disease or Syndrome		Diarrhea, Abdominal pain	Crohn Disease
TIPSS	C0339897	19	Therapeutic or Preventive Procedure	Transjugular intrahepatic portosystemic shunt procedure		
Tongue	C0040408	19	Body Part, Organ, or Organ Component		Mouth problem	Dentistry
Neural	C0599851	18	Functional Concept	Nervous - anatomy qualifier		Nerves System
Inflammatory bowel disease	C0021390	18	Disease or Syndrome	Inflammatory Bowel Diseases	Diarrhea, Blood From Gastrointestinal Tract, Lower/Hematochezia	Inflammatory Bowel Diseases
Clinical Skills	C0008973	18	Qualitative Concept	Clinical Skill		
Metabolism	C0025519	18	Organism Function			Metabolism
Endoscopic Retrograde Cholangiopancreatography	C0008310	17	Diagnostic Procedure			Diagnostic Imaging
Lipids	C0023779	16	Lipid		Lipids Abnormal, Serum	Biochemistry
Sphincter	C1409894	16	Body Part, Organ, or Organ Component			

Tongue	C0040408	16	Body Part, Organ, or Organ Component		Mouth problem	Dentistry
Dysphagia	C0011168	15	Disease or Syndrome	Deglutition Disorders	Dysphagia	
Gastroesophageal reflux disease	C0017168	15	Disease or Syndrome			Gastroesophageal reflux disease
Vomiting	C0042963	13	Sign or Symptom		Vomiting	
Gastric Emptying	C0017127	13	Organ or Tissue Function			
Osteoporosis	C0029456	13	Disease or Syndrome		Osteoporosis	Osteoporosis
Maxillary	C0024947	12	Body Part, Organ, or Organ Component	Maxilla		Musculoskeletal System
Cholecystitis, Acute	C0149520	12	Disease or Syndrome		Acute Abdominal pain, Chronic Abdominal Pain, Vomiting/ Nausea	Cholecystitis
VESICLES	C0333262	12	Acquired Abnormality	Vesicle (morphologic abnormality)		
Liver biopsy	C0193388	10	Diagnostic Procedure	Biopsy of liver (procedure)		
Hepatitis	C0019158	10	Disease or Syndrome		Hepatomegaly, Acute Abdominal Pain, Jaundice, Liver function	Hepatitis

					tests abnormal/serum	
Calcium	C0006675	9	Biologically Active Substance,Element, Ion, or Isotope,Pharmacologic Substance		Hypercalcemia, Hypocalcemia	Biochemistry
disease prevention	C0679698	8	Therapeutic or Preventive Procedure	prevention of disorder		
Lipid	C0023779	7	Lipid	Lipids	Lipids Abnormal, Serum	Biochemistry
Family medicine	C0015607	7	Biomedical Occupation or Discipline	family medicine (field)		Family medicine
GI surgery	C0524722	6	Therapeutic or Preventive Procedure	Gastrointestinal Surgical Procedure		General Surgery
Fluoride Varnishes	C0016326	6	Biomedical or Dental Material,Inorganic Chemical			
Traumatic ulcer	C0406226	5	Injury or Poisoning	Traumatic skin ulcer	Trauma	Trauma
Prevention	C0199176	4	Therapeutic or Preventive Procedure	Prophylactic treatment		
disease prevention	C0679698	3	Therapeutic or Preventive Procedure	prevention of disorder		
fibroma	C0016045	3	Neoplastic Process			Oncology
Caries risk	C1317169	2	Organism Attribute			Dentistry

Concepts	**Concept CUIs**	***Human Development Unit***	**Concept Category**	**Concept Descrition**	**MCC**	**One45**
Voiding	C0042034	79	Organism Function	Urination	Scrotal Mass, Urinary Tract Injuries	Urinary System
Down Syndrome	C0013080	78	Congenital Abnormality,Disease or Syndrome		Genetics Concern	Down Syndrome
Erection	C0030847	75	Organ or Tissue Function	Penile Erection	Impotence/Erectile Dysfunction	Reproductive System
Cystic Fibrosis	C0010674	70	Disease or Syndrome		Blood in Sputum (Hemoptysis/Prevention of Lung Cancer), Cyanosis/Hypoxemia/Hypoxia, Genetic concerns	Cystic Fibrosis
Defect, Neural Tube	C0027794	60	Finding	Neural Tube Defects		Pediatric Neurology
Infections, Respiratory	C0035243	58	Disease or Syndrome	Respiratory Tract Infections	Sore Throat (Rhinorrhea)	Pulmonary Medicine
hereditary non-polyposis colon cancer	C1333990	57	Neoplastic Process	Hereditary Non-Polyposis Colon Cancer Type 1		Colon Cancer

Cystic Fibrosis	C0010674	53	Disease or Syndrome		Blood in Sputum (Hemoptysis/Prevention of Lung Cancer), Cyanosis/Hypoxemia/Hypoxia, Genetic concerns	Cystic Fibrosis
PELVIS	C0559769	48	Body Space or Junction	Pelvic cavity structure	Pelvic mass, Pelvic pain, Prolapse/ Pelvic relaxation	Musculoskeletal System
Genitalia	C0017420	48	Body Part, Organ, or Organ Component		Ambiguous Genitalia	Reproductive System
Female external genitalia	C0227747	43	Body Part, Organ, or Organ Component	Female external genitalia structure	Amenorrhea/Oligomenorrhea	Reproductive System
Hypothalamic	C0020663	42	Body Part, Organ, or Organ Component	Hypothalamic structure	Infertility, Amenorrhea/Oligomenorrhea, Sexual Maturation, Abnormal	Endocrine System
Malformation	C0000768	42	Congenital Abnormality	Congenital Abnormality		
Mosaicism	C0392053	39	Cell or Molecular Dysfunction	Embryonic Mosaic		Embryology

Mutation	C0026882	37	Genetic Function		Genetic Concern	Medical Genetics
Cri du chat	C0234861	36	Sign or Symptom		Genetic Concern	Medical Genetics
Sexuality	C0036864	36	Behavior	Sex Behavior		
Cancer	C0006826	36	Neoplastic Process	Malignant Neoplasms	Prostate cancer, Prevention of lung cancer	Oncology
Labor	C0022864	35	Organism Function	Labor (Childbirth)	Antepartum care, Intrapartum Care/Postpartum Care	obstetric
Pregnancy	C0032961	34	Organism Function		Pregnancy	Obstetrics and Gynecology
Placenta	C0032043	33	Embryonic Structure		Failure To Thrive, Infant/Child, Non-Reassuring Fetal Status (Fetal Distress), Pelvic pain, Prematurity	Embryology
Phenylalanine	C0031453	33	Amino Acid, Peptide, or Protein,Biologically Active Substance,Pharmacologic Substance			Biochemistry

Neoplasia	C0027651	31	Neoplastic Process	Neoplasm		Oncology
Prostate gland	C0033572	31	Body Part, Organ, or Organ Component	Prostate	Impotence/Erectile Dysfunction,	Endocrine System
Reproductive	C0035150	29	Organism Function	Reproduction		Reproduction System
Huntington Disease	C0020179	27	Disease or Syndrome			
Fetal	C0015965	27	Embryonic Structure	Fetus	Non-Reassuring Fetal Status (Fetal Distress)	Embryology
Sexual Health	C2362326	24	Intellectual Product	Sexual Health Topics		
Testis	C0039597	22	Body Part, Organ, or Organ Component		Scrotal Mass, Scrotal pain	Endocrine System
Sexuality	C0036864	19	Behavior	Sex Behavior		
Breast	C0006141	19	Body Part, Organ, or Organ Component		Breast Disorders, Breast Lump/Screening	
Global Health	C1456573	18	Group Attribute			

AML	C0023467	17	Neoplastic Process	Leukemia, Myelocytic, Acute		Leukemia
Chromosomes	C0008633	16	Cell Component		Genetic Concern	Medical Genetics
Prenatal	C0033052	15	Health Care Activity	Prenatal care	Antepartum Care	Prenatal care
PKU	C0031485	15	Disease or Syndrome	Phenylketonurias		
Vagina	C0042232	15	Body Part, Organ, or Organ Component		Vaginal Bleeding, Excessive/Irregular/Abnormal, Vaginal Discharge	Reproductive System
contraceptive	C0009871	14	Pharmacologic Substance	Contraceptive Agents	Contraception	Contraception Request
Human Development	C0020119	14	Physiologic Function			Physiology
IVF	C0015915	13	Therapeutic or Preventive Procedure	Fertilization in Vitro	Infertility	Infertility
Anatomy	C0002808	13	Biomedical Occupation or Discipline	Science of Anatomy		

Histology	C0019638	10	Biomedical Occupation or Discipline			Histology
Health Policy	C0018735	10	Regulation or Law			
Reproductive	C0035150	8	Organism Function	Reproduction		Reproductive System
Menstrual cycle	C0025329	5	Organism Function		Menstrual Cycle, Abnormal	Reproductive System
ED	C0242350	4	Disease or Syndrome	Erectile dysfunction	Impotence/ Erectile Dysfunction	

Concepts	Concept CUIs	ProComp Unit	Concept Category	Concept Descrition	MCC	One45
Research	C0242481	86	Research Activity	Research Activities		
public health	C0699943	90	Health Care Activity	Public health service	Population Health	
HIV	C0019693	111	Disease or Syndrome	HIV Infections	Vaginal Discharge/Vulvar Itch/STD,	Human Immunodeficiency virus
Blood	C0005767	88	Tissue		Blood From Gastrointestinal Tract, Blood in urine, Blood in sputum.	Hemic System

Concepts	Concept CUIs	Clinical Skill Objective Unit	Concept Category	Concept Descrition	MCC	One45
General physical examination	C0031809	89	Health Care Activity	Physical Examination		