Poster presented at the International Conference on Biomedical Ontology (ICBO) 2014.

Coverage of Rare Disease Names in Clinical Coding Systems and Ontologies and Implications for Electronic Health Records-Based Research

Rachel Richesson
Duke University School of Nursing
Durham, NC USA
rachel.richesson@dm.duke.edu

Kin Wah Fung
National Library of Medicine
Bethesda, MD, USA

Olivier Bodenreider
National Library of Medicine
Bethesda, MD, USA

***Abstract*—This poster will present the completeness of coverage of rare disease names in standard coding systems, including the International Classification of Diseases (ICD) and SNOMED CT, and ontologies such as the Orphanet Rare Diseases Ontology (RDO). Using use cases and a set of 45 rare diseases for the national Patient Centered Outcomes Research Network (PCORnet), the poster will describe the current capacity and implications for electronic health records-based research on these diseases. Authors will provide suggestions on how clinical coding systems and ontologies can be used in a coordinated approach to support the use of electronic health record data for various types of research related to rare diseases.**

Keywords—rare diseases; clinical classifications; ontologies; biomedical research; electronic health records

I. Introduction

Rare diseases are defined in the US as conditions that affect less than 200,000 Americans and in the European Union as those with a prevalence of 5 per 10,000 or less.[1,2] The NIH Office of Rare Diseases Research recognizes 6,485 rare diseases.[3] Although each rare disease is uncommon, collectively they constitute a significant burden to the health care system. One estimate suggests that 1 in 10 Americans are affected by a rare disease.[2] Consequently 'rare diseases' have emerged as priority topics in public health and research. Rare disease names are included, at different levels of completeness and granularity, in a number of clinical coding systems that are embedded in electronic health record (EHR) systems, and in a number of ontologies designed to support the diagnosis rare diseases and investigation of their causes and treatments.[4]

With increased adoption and "meaningful use" of EHRs, there is renewed effort in leveraging EHRs for research. In the U.S., the national Patient Centered Outcomes Research Network (PCORnet) was funded this year from the Affordable Care Act to examine real-world treatment decisions, and is specifically tasked to conduct observational and interventional research on the comparative effectiveness of various treatments, using distributed and heterogeneous EHR systems.[5] The PCORnet research portfolio currently includes 45 rare diseases (in addition to approximately 20 more common conditions). The objective of this poster is to determine the coverage of these rare disease names in standard coding systems and explore the current capacity and implications for EHR-based research on these and other rare diseases.

II. Methods

In this poster we present an inventory of various clinical coding systems and ontologies that are relevant to rare diseases, and summarize their coverage of rare disease names from previous studies. We match rare diseases names and synonyms from the Office of Rare Disease Research (ORD) and Orphanet (RDO) to the Unified Medical Language System (UMLS) Metathesaurus and identify maps to SNOMED CT and other terminologies. To characterize the coverage of rare diseases studied in PCORnet, we estimate the number of precise and equivalent matches in the three clinical classifications (ICD-9-CM, ICD-10-CM, and SNOMED CT) for a set of 45 rare diseases studied in PCORnet. Finally, we present the likely use of existing classifications, ontologies, mappings, and tools to support the research process, from the collection of data in clinical settings to their use in various types of EHR-based research.

III. RESULTS

SNOMED CT has the highest coverage of rare disease names among clinical terminologies in UMLS, and covers 44% of the 6,485 diseases (19,504 terms) recognized by the Office of Rare Diseases (ORD), and 48% of the 6,750 diseases (15,585

terms) diseases listed in the Orphanet Rare Disease Ontology. 25% (1,611) of ORD and 14% (1,592) RDO disease names have bi-directional one-to-one maps to SNOMED CT. The rest are one-to-many or many-to-one maps. Two terminologies have higher coverage than SNOMED CT. Medical Subject Headings (MeSH) covers 75% and 70%, while Online Mendelian Inheritance in Man (OMIM) covers 49% and 57%, of ORD and RDO respectively. Overall, the UMLS covers 82% of ORD-recognized and 84% of RDO-recognized rare diseases.

All of the rare diseases studied in PCORnet were included in the UMLS and its source terminologies. 8 diseases did not have any match to SNOMED CT, ICD-9-CM or ICD-10-CM. The 45 rare diseases studied in PCORnet yielded multiple matches to terms in clinical coding systems; i.e., many PCORnet rare disease names matched to more than one (term) code in a coding system, and many codes from clinical coding systems matched more than one rare disease name. Of 55 ICD-9-CM codes that matched to a PCORnet rare disease, 7 were matched to multiple rare diseases. Of 47 matched ICD-10-CM codes, 4 matched to multiple rare diseases, and of 59 matched SNOMED CT codes, one SNOMED CT code matched to multiple PCORnet rare diseases. The proportions of matched codes that were considered equivalent matches (rather than broader matches or related terms) were 25%, 45% and 94% for ICD-9-CM, ICD-10-CM and SNOMED CT respectively.

IV. Conclusions

The coverage and quality (i.e., precision and equivalence) of terms for rare diseases in clinical coding systems is less than ideal, but is markedly improved with SNOMED CT in comparison to ICD 9 and 10 classifications. The lack of precise and complete coverage of rare disease names in clinical coding systems will inhibit the automated identification patients with rare diseases from EHR data for clinical trial recruitment or observational research. The coverage of rare disease names in specialized ontologies (e.g., OMIM) is higher, but these are not designed for use in clinical EHR systems.

Given the intended purpose for each classification and ontology and the completeness and coverage of rare disease names, we propose how these various clinical coding systems, ontologies, and UMLS mappings can be leveraged to support an efficient national research infrastructure and learning healthcare system. The UMLS is a vital tool to support the linkage across clinical coding systems and specialized ontologies that will be essential for a national EHR-based rare diseases research infrastructure.

Ontologies can support advances in understanding disease etiology and potential treatments. Specialized ontologies, such as OMIM, RDO, and others (such as the Human Phenotype Ontology) can provide the vocabulary for detailed clinical documentation , or “deep phenotyping”, of genetic diseases (e.g., in the NIH Undiagnosed Diseases Network), and complement clinical terminologies and administrative classifications widely used in EHRs. This poster will include an illustrative representation of the collection of rare disease-specific data in dedicated ontologies to support diagnosis, and the use of mappings to standardized clinical terminologies or classifications as needed for clinical documentation, data exchange, billing and public health reporting.

Acknowledgment

This work was partly supported by the Intramural Research Program of the National Institutes of Health and the National Library of Medicine. This work was also supported in part by PCORnet, funded by the Patient Centered Outcomes Research Institute (PCORI).

References

[1] Orphanet. *Orphanet Rare Disease Ontology (ORDO)*. 2014 [cited 2014 March 14]; Available from: http://www.orphadata.org/cgi-bin/inc/ordo_orphanet.inc.php.

[2] NORD. *Rare Disease Information*. 2014 [cited 2014 March 14]; Available from: http://www.rarediseases.org/rare-disease-information.

[3] NIH. *Office of Rare Diseases Research (ORDR) Brochure*. 2009 [cited 2010 20/08/2010]; Available from: http://rarediseases.info.nih.gov/asp/resources/ord_brochure.html.

[4] Fung, K.W., R.L. Richesson, and O. Bodenreider, Coverage of Rare Disease Names in Standard Terminologies and Implications for Patients, Providers, and Research, in Paper accepted for presentation: American Medical Informatics Association Annual Symposium2014: Washington, D.C.

[5] PCORI. *PCORnet: The National Patient-Centered Clinical Research Network*. 2014 [cited 2014 March 13]; Available from: http://www.pcori.org/funding-opportunities/pcornet-national-patient-centered-clinical-research-network/.