INTERNATIONAL HEALTH TERMINOLOGY STANDARDS DEVELOPMENT ORGANISATION

Coverage of human phenotypes in SNOMED CT

Presenter: Rainer Winnenburg/Olivier Bodenreider, National Library of Medicine

Audience

Researchers involved with EHR/clinical data warehouse data in support of translational research.

Objectives

To assess the coverage of the human phenotypes in SNOMED CT.

Abstract

Background: While the past decades have seen unprecedented efforts directed towards genotyping, parallel efforts are required on the side of phenotyping in order to understand how genetic variation relates to clinical manifestations. The potential of using coarse phenotypes (at the disease level) based on electronic health record (EHR) data for genomic studies has been demonstrated. However, the study of rare syndromes will likely require detailed phenotyping. Specialized ontologies of phenotypic abnormalities, such as the Human Phenotype Ontology (HPO)[1], are currently used for the annotation of phenotypes in databases. However, the use of EHR data for translational research is predicated on the analysis of phenotypes coded natively with terminologies such as SNOMED CT, and therefore on the coverage of such phenotypes in SNOMED CT.

Methods: We map HPO terms (April 2014) to the Unified Medical Language System (UMLS) Metathesaurus (2013AB) using exact and normalized matches, followed by semantic filtering. Then we map HPO classes to SNOMED CT concepts (July 2013) through UMLS concepts. We also compare these lexical mappings to the cross-references to SNOMED CT provided by HPO.

Results: We extracted for 10,491 HPO phenotypes (classes), their preferred terms (one per class) and 5,923 synonyms, along with the cross-references to SNOMED CT provided by HPO for 267 classes. In total, some cross-reference or lexical mapping to UMLS was found for 5,858 HPO classes (56%). Using the lexical mappings to UMLS obtained for 5,650 HPO classes (54%), we derived mappings to SNOMED CT for 3,116 classes (30%). Although coverage is partial, SNOMED CT is the source vocabulary in the UMLS for which the coverage of HPO phenotypes is the highest. HPO only provides cross-references to SNOMED CT for 3% of its classes, which is in contrast to the Medical Subject Headings (MeSH), to which most cross references are provided (covering 10% of the classes).

Conclusions: Although our mapping has not been validated, this investigation suggests that our approach to mapping HPO classes to SNOMED CT through the UMLS yields significantly more links between HPO and SNOMED CT than are provided in HPO through cross-references. This approach can therefore facilitate the interoperability between these two terminological resources. However, harmonization between HPO and SNOMED CT is needed to ensure that detailed phenotypic information can be captured in clinical repositories.

References

1 Kohler, S., et al. (2014) The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data, Nucleic Acids Res, 42, D966-974.