Mondo Workshop April 2021
Report

New view of the Mondo Disease Ontology High-level Classification According to Harrison's Principles of Internal Medicine textbook.

Authors:
Sabrina Toro[1], Ada Hamosh[2], Nico Matentzoglu[3], Monica Munoz-Torres[1], Lauren Chan[4], Courtney Thaxton[5], Gioconda Alyea[6], Melissa Haendel[1], Chris Mungall[7], Peter N Robinson[8], Nicole Vasilevsky[1]

1. University of Colorado Denver, Anschutz Medical Campus, Aurora, CO USA
2. Johns Hopkins University, Baltimore, MD, USA
3. Semanticly, Athens, Greece
4. Oregon State University, College of Public Health and Human Sciences, Corvallis, OR USA
5. University of North Carolina, Chapel Hill, NC, USA
6. National Center for Advancing Translational Sciences, National Institute of Health, Bethesda, MD, USA
7. Lawrence Berkeley National Laboratory, Berkeley, CA, USA
8. Jackson Laboratory, Farmington, CT, USA

Introduction

The Mondo Disease Ontology (Mondo) integrates multiple disease terminologies into a coherent logic-based ontology that provides precise semantic mappings between terms [1]. It harmonizes several source terminologies and ontologies covering various aspects of disease across species, including the Online Mendelian Inheritance in Man (OMIM) (Mendelian disease) [2], Orphanet (rare diseases) [3], the neoplasm branch of the National Cancer Institute Thesaurus (NCIt) [4], GARD (rare diseases) [5], Disease Ontology (human diseases) [6], MedGen (genetic diseases) [7], and broad terminologies such as Medical Subject Headings (MeSH) [8], SNOMED Clinical Terminology [9], International Classification of Disease (ICD) [10], ICD-Oncology (ICD-O) [11], and numerous others.

Many groups use Mondo for their research and clinical ontologies, databases, and portals (https://mondo.monarchinitiative.org/pages/users/). The Mondo team welcomes feedback from the community, including medical experts, to improve the ontology and ensure that it correctly represents diseases based on the current knowledge and that it also remains relevant and useful to this wide variety of users. Mondo provides multiple ways to classify a disease based on a variety of criteria (such as the etiology, the origin, the anatomical structure affected, etc., of the disease), and its hierarchical classification allows for multi-inheritance, whereby terms can be classified under more than one parent. While this classification is useful for computational purposes, a simplified and more practical representation may be preferred for some users [12,13].

Medical and ontology experts met at a workshop in April 2021 to investigate a new classification that may complement/correspond with the clinical community's approach to disease classification. The “Harrison's Principles of Internal Medicine” textbook (below called ‘Harrison’) has been used as a landmark in medicine, for both trainees and practicing clinicians[14]. The goal of this workshop was to review the high-level classification of Mondo to reflect the organization of Harrison's book of medicine, and thereby offer a clinically oriented view of Mondo (called “Harrison view”). This report summarizes the changes made to the Mondo ontology and the creation of a new “Harrison view” of Mondo. The Mondo high-level classification before and after this work, as well as the “Harrison view” of Mondo are shown in Figure 1. Details of the work can be viewed on GitHub (https://github.com/monarch-initiative/mondo/issues?q=is%3Aissue+label%3A%22Harrison+reclassification%22+).

Updates to the Mondo Ontology

We compared the Mondo classification to the organization of medical knowledge in Harrison’s. We identified the Mondo high-level terms to be removed from the high-level classification, new high-level grouping terms to be created, and high-level terms that should remain in Mondo but would not be represented in the Harrison classification. Mondo terms reflecting the Harrison classification were tagged for the generation of the “Harrison view”.

Terms removed from the Mondo high-level classification

Some Mondo terms were clearly too specific to be high-level classes, such as ‘serpinopathy’ (MONDO:0027749), ‘segmental odontomaxillary dysplasia’ (MONDO:0019029), and ‘regional odontodysplasia’ (MONDO:0019367). These terms were reclassified as children of other existing terms, Mendelian disease (MONDO:0003847), bone development disease (MONDO:0005497), and periodontal disease (MONDO:0002635), respectively, and therefore removed from the Mondo high-level classification.

‘Disorder involving pain’ (MONDO:0021668) was defined as a grouping class for diseases for which “pain” was the main feature (Equivalent Axiom: `'disease or disorder' and ('disease has major feature' some Pain)`). Pain is a feature of many diseases and it could be arbitrary whether it is considered as the main feature, therefore this term was not precise to be a useful high-level term. Children terms of this class, such as ‘chronic pain syndrome’ (MONDO:0024317) and ‘headache disorder’ (MONDO:0021146) referred to ‘neurological pain disorder’, nervous system disorders that have pain as a major feature. A new grouping class with that name was created (MONDO:0700057) under the parent term ‘nervous system disorder’ (MONDO:0005071), and ‘disorder involving pain’ (MONDO:0021668) was obsoleted.

The term ‘radiation or chemically induced disorder’ (MONDO:0045028) was a grouping term for ‘chemically-induced disorder’ (MONDO:0029001) and ‘radiation-induced disorder‘

(MONDO:0043459). This grouping term was uninformative and deemed an unnecessary grouping class and therefore was obsoleted. Similarly, ‘chemically-induced disorder’ (MONDO:0029001) had only one asserted child, ‘poisoning’, and was not an informative grouping class. The term, ‘poisoning’ (MONDO:0029000) was made a high-level class in Mondo, similar to the Harrison classification. The grouping class ‘radiation-induced disorder‘ (MONDO:0043459) is useful to some users, therefore we kept this term and made it a Mondo high-level class. Terms subsumed by this class were reviewed, and the superclass assertion to 'radiation-induced disorder' was deleted, such that all children terms of ‘radiation-induced disorder‘ were inferred (i.e. not asserted) based on the class Equivalent Axiom `('disease or disorder' and ('realized in response to' some 'exposure to electromagnetic radiation'))`. However, ‘radiation-induced disorder‘ is not represented in the Harrison organization, and therefore this term was not added to the “Harrison view”.

The Harrison organization distinguishes between ‘nutritional disease’ and ‘metabolic disease’. Therefore we obsoleted the grouping class ‘nutritional or metabolic disease’ (MONDO:0024297), and made its children terms ‘nutritional disorder’ (MONDO:0005137) and ‘metabolic disease’ (MONDO:0005066) high-level classes with the “Harrison view” tag.

New high-level Mondo classes included in the “Harrison view”.
In addition to the Mondo terms made high-level classes following the obsoletion of their grouping terms as described above, a new term was created to group existing high-level classes: the Mondo high-level classes ‘pregnancy disorder’ (MONDO:0024575) and ‘puerperal disorder’ (MONDO:0044013) were grouped under the more general high-level term ‘obstetric disorder’ (MONDO:0700003), representing disorders associated with pregnancy, childbirth, and puerperium (the period following childbirth during which the mother's reproductive organs return to their nonpregnant condition).

Existing high-level Mondo classes included in the “Harrison view”:
The term ‘endocrine system disease’ (MONDO:0005151) was added to the “Harrison view” as a high-level term. Additionally, 'cell proliferation disorder' (MONDO:0045024) was also added to the “Harrison view”, but was renamed ‘cancer or benign tumor’, a designation more familiar to clinical users.

Existing high-level Mondo classes not included in the “Harrison view”:
Some high-level Mondo terms may not be clinically relevant but are still useful to other Mondo users, and therefore were kept as high-level terms in the ontology. These terms are still found in the ontology, but are not included in the “Harrison view”. These high-level classes include ‘non-human animal disease’ (MONDO:0005583) and high-level classification created based on axioms, such as ‘disease by anatomical system’ (MONDO:0021199), ‘radiation-induced disorder‘ (MONDO:0043459, see above), and the newly created class ‘idiopathic disease’ (MONDO:0700007).

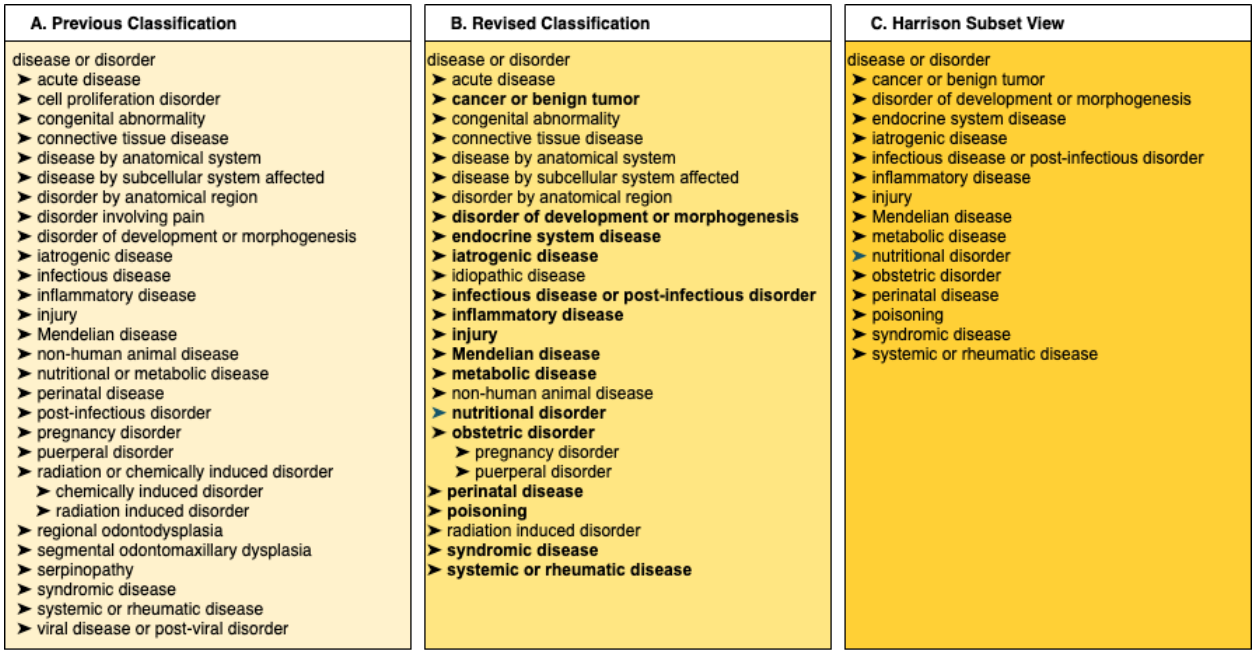

Figure 1: Top-level hierarchy in Mondo. **A.** The previous classification shows the top-level classes before our revision efforts. **B.** The revised classification shows the current top-level classes. The bold terms are in the "Harrison view" subset. **C.** The Harrison subset view (referred to as "Harrison view") of the top-level classes in Mondo.

Comparison of the Mondo classification and the "Harrison view":

The Mondo high-level classification and the "Harrison view" are shown in Figure 1. The "Harrison view" includes fewer high-level terms. The number of total terms available via this view is 19380, which is 87% of the total number of terms available in Mondo. 2% of the terms excluded from the human disease-centric "Harrison view" refer to non-human animal diseases. Unexpectedly, some 'non-human animal disease' terms (59 terms) were found in the "Harrison View", most of them being automatically classified as 'infectious disease' (MONDO:0005550, present in the "Harrison View") based on logical axioms specifying the infectious agent. 7% of the terms excluded from the "Harrison view" are terms referring to 'disease susceptibility'. These excluded terms are solely classified as 'disease susceptibility'; other terms, such as 'breast-ovarian cancer, familial, susceptibility to, 1' (MONDO:0011450) are present in the "Harrison View" based on their classification to another parent, in this case 'Mendelian disease' (MONDO:0003847). 90% of the terms excluded from the "Harrison view" are diseases classified solely under the classes 'disease by anatomical system' (MONDO:0021199), 'disorder by anatomical region' (MONDO:0024505), and/or 'disease by subcellular system affected' (MONDO:0021194). The majority of these terms may still be valuable and useful to the clinical community, for example 'psychiatric disorder' (MONDO:0002025) and its children, some 'nervous system disorder' (MONDO:0005071) terms, and some 'cardiovascular disease' (MONDO:0004995) terms.

Future work:

The majority of the terms excluded from the “Harrison View” are defined by the anatomical system (UBERON:0000467) and region (UBERON:0000475) where the features of the disease are found, and/or the subcellular structure, the cellular process, and/or the molecular activity affected in the disease. We will review these terms and investigate whether they can be included in an existing “Harrison View” class, or be established as a new high-level class in the “Harrison View” (for example, ‘psychiatric disorder’-MONDO:0002025).

The ‘non-human animal disease’ terms present in the “Harrison View” will be reviewed to ensure that they are not included in this human disease centric view. This review might include changes in the term parentage and adjustment of existing axioms.

Summary:

We created the “Harrison view” to make Mondo more accessible for clinicians. This view follows a classification comparable to the organization of the Harrison's textbook of medicine, familiar to clinicians, and will be available in the Monarch instance of the Ontology Lookup Service (OLS) (https://ols.monarchinitiative.org/index) and for download in upcoming releases, which are available at https://github.com/monarch-initiative/mondo/releases. This work was also the opportunity to review the high-level classification of Mondo, update existing terms, and create new ones. The Mondo ontology files are available for download at https://github.com/monarch-initiative/mondo ; feedback can be submitted on our GitHub issue tracker here: https://github.com/monarch-initiative/mondo/issues

References:

1. Vasilevsky, N. *et al.* Mondo Disease Ontology: harmonizing disease concepts across the world. in *CEUR Workshop Proceedings* vol. 2807 (CEUR-WS, 2020).
2. Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). https://omim.org/ (2007).
3. Orphanet: an online database of rare diseases and orphan drugs. Copyright, INSERM 1997. http://www.orpha.net.
4. Sioutos, N. *et al.* NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information. *J. Biomed. Inform.* **40**, 30–43 (2007).
5. Lewis, J., Snyder, M. & Hyatt-Knorr, H. Marking 15 years of the Genetic and Rare Diseases Information Center. http://dx.doi.org/10.3233/trd-170011 (2017) doi:10.3233/trd-170011.
6. Schriml, L. M. *et al.* Human Disease Ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Res.* **47**, D955–D962 (2019).
7. Louden, D. N. MedGen: NCBI’s Portal to Information on Medical Conditions with a Genetic Component. *Med. Ref. Serv. Q.* **39**, 183–191 (2020).
8. *Medical Subject Headings (MeSH)* https://www.ncbi.nlm.nih.gov/mesh.
9. *SNOMED International* https://www.snomed.org/.
10. World Health Organization. ICD-10 : international statistical classification of diseases and related health problems : tenth revision. Spanish version, 1st edition published by PAHO as Publicación Científica 544 (2004).
11. World Health Organization. International classification of diseases for oncology (ICD-O) – 3rd edition, 1st revision. viii, 242 p. (2013).
12. Chute, C. G. The rendering of human phenotype and rare diseases in ICD-11. *J. Inherit.*

Metab. Dis. **41**, 563–569 (2018).
13. Cornet, R. & Chute, C. G. Health Concept and Knowledge Management: Twenty-five Years of Evolution. *Yearb. Med. Inform.* **Suppl 1**, S32–41 (2016).
14. Larry Jameson, J. *et al. Harrison's Principles of Internal Medicine, Twentieth Edition (Vol.1 & Vol.2)*. (McGraw-Hill Education, 2018).