Special Lecture | JCMI/APAMI Special Keynote Session | ICD11

[JCMI/APAMI Special Keynote Session 3] ICD11

Sun. Nov 22, 2020 11:20 AM - 12:20 PM Hall-A (Middle Hall)

[SKS-3-01] The Potential Roles of ICD-11 in Translational Research: Architectural Features that Support Data Science

*Christopher G. Chute[1] （1. Professor of Medicine, Public Health and Nursing, Johns Hopkins University, USA）

Keywords: ICD-11, Architecture of the Classification, Data Science

The International Classification of Disease (ICD) has progressed from its mid-nineteenth century origins as a short list of causes of death, to become the dominant classification of human diseases, syndromes, and conditions around the world. Despite this profound increase in content and impact over more than a century between the initial versions of the International Statistical List and the release of ICD-10 in 1990, there as virtually no evolution of the structure or architecture of the classification. It remained essentially a table of terms with associated code values and very little more. Some have characterized the ICD as a 16th century spreadsheet, harkening to the structure of the ICD’ s ancient predecessor, the London Bills of Mortality established during the reign of Henry VIII of England.

The Potential Roles of ICD-11 in Translational Research: Architectural Features that Support Data Science

Christopher G Chute[a]

[a] *Bloomberg Distinguished Professor of Health Informatics, Professor of Medicine, Johns Hopkins University, USA*

The International Classification of Disease (ICD) has progressed from its mid-nineteenth century origins as a short list of causes of death, to become the dominant classification of human diseases, syndromes, and conditions around the world. Despite this profound increase in content and impact over more than a century between the initial versions of the International Statistical List and the release of ICD-10 in 1990, there as virtually no evolution of the structure or architecture of the classification. It remained essentially a table of terms with associated code values and very little more. Some have characterized the ICD as a 16th century spreadsheet, harkening to the structure of the ICD's ancient predecessor, the London Bills of Mortality established during the reign of Henry VIII of England.

The ICD11 Opportunity

The leadership of the Classifications and Terminology team at the World Health Organization (WHO) were very much aware of the shortcoming of the ICD in the modern digital age. Around 2005 they organized meetings to explore how ICD could be revised to leverage the advances in computer science, ontology, and knowledge representation that had accelerated through the 20th and early 21st centuries. From its inception, the next revision of the ICD, ICD11, was intended to leap-frog classification tradition and embrace the modern digital revolution. The only question was how to reconcile that brave new vision with the traditional needs and requirements of the statistical mortality and public health communities who had deep dependencies on the centuries of evolved structure.

ICD11 was developed by many teams of clinical specialists and domain leaders in a series of 19 topic advisory groups (TAGs), with the Internal Medicine TAG in turn having 8 subspecialty groups. Each of these TAGs and groups comprised between a dozen to more than a score of experts, collaborating to fundamentally revise the science and knowledge base of ICD11. All of the TAGs reported to a Revision Steering Group (RSG), comprised of the TAG chairs and the RSG chair. However, there was a special TAG for Informatics, which contributed greatly to the ICD11 architectural development.

Architecture of the Classification

The Content Model

The core content model of the ICD expands greatly upon the simple historical term and code structure in historical ICD versions. Each term or concept in ICD has these potential elements:

1. Concept title
2. Unique Identifier (URI)
3. Fully specified name
4. Synonyms
5. Classification Properties
6. Parent and child relationships
7. Brief definition
8. Long description
9. Body system
10. Manifestations
11. Etiology
12. Genomic association
13. Causal Agents
14. Severity grades
15. Temporality
16. Functional impact

The first eight are required and fully populated in the released version. The remaining elements are variously complete, and remain an opportunity for future work.

Foundation and Linearizations

All concepts in the ICD are rendered in the Foundation component, which is an acyclic graph of all concepts and their relationship trees. Unlike the historical ICDs, the Foundation may have multiple inheritance, where a single term may have one or more, sometimes many more, conceptual parents. Thus, stomach cancer can be a child in the cancer tree has it has been for decades, but can also be a child in the gastro-intestinal illness chapter where is was previously absent. The allows for the assertion of ontological structures and supports complex navigation in the Foundation, previously impractical in earlier versions. The depth of this semantic network is virtually unlimited, meaning highly specialized conceptual children can exist in this network without any limitation of digits in a coding structure. It can be vastly larger than any historical version of the ICD, since this is not the system used for practical disease coding. It is the conceptual underpinning of the entire system.

The Foundation does violate a key precepts of statistical classification, which are they the content must be mutually exclusive and exhaustive. Mutually exclusive concepts in practice mean that they must have only one place in a concept hierarchy, and must therefore have only one parent term. Exhaustive classifications are achieved by addition residual categories such as “other” or “not specified” at the terminus of concept branches. This is not the architecture of the Foundation, so we developed Linearizations derived from the Foundation that would have these properties. These Linearizations are effectively a strict walking of the Foundation concept tree to a limited depth, and deliberately choosing a single parent for inheritance; they thus satisfy the requirements of being mutually exclusive and exhaustive. In an interesting extension, ICD11 can support multiple, simultaneous linearizations from the Foundation. These include the main tabular publication of the Morbidity and Mortality Statistics Linearization, as well as general linearizations for Primary Care, or for example a sub-specialty linearization in Dermatology.

Post-coordination

The content of the Foundation provides an enormously rich thesaurus for the ICD, and in fact the functional index of ICD11 is built from the Foundation. However, even greater expressivity can be achieved by the combination of base terms such as a disease entity with qualifier codes. ICD11 contains a complete chapter of such qualifier or extension codes, that can be combined with terms to compose “sentences” of clinical description. Thus a given cancer can be modified to include histology, anatomic site, stage, and extent as a single ICD11 assertion. This compositional structure enables profoundly granular, detailed, and specific description of clinical entities. ICD11 effectively marries the aggregation capacity of a clinical classification with the potential of a highly expressive terminology.

Data Science Implications

Given the deep capacity of ICD11 to capture specific and detailed coding of diseases and conditions, it can provide a framework for clinical research and discovery emerging from encoded medical records. ICD remains the mostly widely used classification in the world. Successful deployments of ICD11 will profoundly benefit societies capacity for detailed observational research, where EHR data can be a source for data science, outcomes research, and epidemiological investigation.

Disclosures: Dr. Chute served as the RSG chair in the ICD11 revision process, and presently co-chairs the Medical Scientific Advisory Committee at WHO for the ICD classification.

Address for correspondence

Christopher G Chute, MD, DrPH
Bloomberg Distinguished Professor of Health Informatics
Professor of Medicine, Public Health, and Nursing
Johns Hopkins University
Baltimore, MD 21287
E-mail: chute@jhu.edu