Review of endpoint/outcome measures ontologies

Criteria used: structure, completeness & accuracy, fitness for purpose

GI_cancer ontology

Summary

Overall this is a fairly accurate but incomplete ontology. It would benefit from more classes being imported from the Cancer Care: Treatment Outcome Ontology (CCTOO), and from some manual curation effort to improve its structure and synonym coverage. Use in Named Entity Recognition (NER) was reasonable – around 50% of relevant entities from a selection of appropriate clinical trial documents were annotated by using the ontology and its synonyms directly.

Structure

- Simple taxonomy structure which is sufficient
- A subset of classes are imported from the Cancer Care: Treatment Outcome Ontology (CCTOO). It doesn't look like this ontology is maintained any longer. This would need to be taken into account when considering maintenance.
- Some very specific classes are placed directly under general classes e.g. "Frequence of positive ctDNA" under "Measure of efficacy". The structure would be improved by adding some intermediate classes.
- Conversely, general classes are sometimes erroneously found under specific classes for example "Pain" (RTEJ4J6zE7DIrv0hhB5nQq) is under "Treatment-related adverse events" (CCTO:000098) when clearly pain is not always an adverse effect of medical treatment.

Completeness and accuracy

- Some things that didn't feel like they belong in a clinical endpoints ontology, such as "Measures of costs" and "Measure of healthcare resource utilization" which are not clinical measurements.
- I observed some duplication of classes that should be combined e.g. Overall Response Rate (CCTO:000068) and Objective response rate (RDFgLwku1LlZiZMgEANIiVQ) mean the same thing.
- Synonyms are provided for a subset of classes. Where present I found these to be appropriate. It would be good to have more synonym coverage generally.
- Names and abbreviations of specific tests and questionnaires were missing e.g. CTCAE v4.0, QLQ-STO22, EORTC QLQ-C30.
- Definitions where present were generally good quality

Fit for purpose

- The key use-case for this ontology is Named Entity Recognition (NER). I tested the coverage by annotating several clinical trial outcomes sections with this ontology, including its synonyms.
- I found the ontology was missing some very common endpoint synonyms like "Quality of Life" or "QoL" for RCM23q2ZjrErzJqN4ot7ko and "Pathologic complete response".
- Some common acronyms were also not included as synonyms e.g. "OS" as a synonym for "Overall Survival"

Diabetes ontology

Summary

This ontology is broadly accurate, and has good detail in some parts e.g. questionnaires. It would be improved by adding more synonyms, and having a subject matter expert check for the presence of critical endpoints.

Structure

- Simple taxonomy structure which is sufficient
- Top-level hierarchy is reasonable
- Class labels are often camel-cased with no spaces which makes it hard to search and read
- Capitalization of class labels should be standardised for consistency
- Acronyms shouldn’t be part of class labels e.g. “Time to Maximal Concentration (TMAX)”
- Some classes only have an acronym and no synonym with the full label e.g. “AUROC” (R9PLgQOvMbUEg44aBV6tIr6) and “DKA rate” (RX6HTkQuUJHVGBXootOHin)

Completeness and accuracy

- Missing some key diabetes measurements like “fasting blood glucose” and “CGM glucose”
- Coverage of questionnaires is good
- Most classes have no synonyms
- Some classes have definitions, and these are mostly accurate although a little informal in places e.g. “Heart Failure” is defined as “means that the heart isn't pumping as well as it should”.
- Some inaccurate classifications observed e.g. “Plasma” is_a “Biomarker”

Fit for purpose:

- The key use-case for this ontology is Named Entity Recognition (NER). I tested the coverage by annotating several clinical trial outcomes sections with this ontology, including its synonyms.
- Because this ontology doesn’t have many synonyms, annotation of clinical trial documents was limited.
- Approx 30% of endpoints were annotated using this ontology and its synonyms directly.