Representation of investigational drugs in the OMOP CDM

Michael Kallfelz, Dmitry Dymshyts, Meera Patel, Christian Reich, Jeremy Warner, Rimma Belenkaya
OHDSI Oncology Workgroup

Background

The OMOP standardized vocabularies are primarily based on terminologies in use for standard patient treatment. To date, whenever investigational drugs or treatments are in focus of the analysis, these must be created as local concepts. In order to support researchers transforming their primary source data to meaningful OMOP concepts, a dedicated vocabulary for investigational drugs is required. As a first step, possible sources were collected and assessed for usability in determining investigational versus regularly approved medications. One particular question was to enable search for preliminary manufacturer designations for the substances in question (e.g. AZD 6094 for Savolitinib) as these are frequently present in source data. A challenge in this approach are proprietary designations that are not publicly known, therefore limiting access to data containing these. Use cases requiring a representation of investigational drugs are currently ongoing research within clinical datasets as well as the re-use of existing historic Clinical Trial datasets focused on drug treatment.

Methods

Investigation was carried out to identify possible reliable and free sources by searching for keywords indicating an investigational use of drugs, primarily in an Oncology setting. DrugBank[1] and the NCI Thesaurus (NCIt)[2] were identified as sources that in combination provide adequate coverage and depth of information as well as meet the requirement to create a vocabulary of human drugs under investigation. Other sources such as PubChem[3] and ChemIDPlus[4] provide excellent content and can be used to fill gaps.

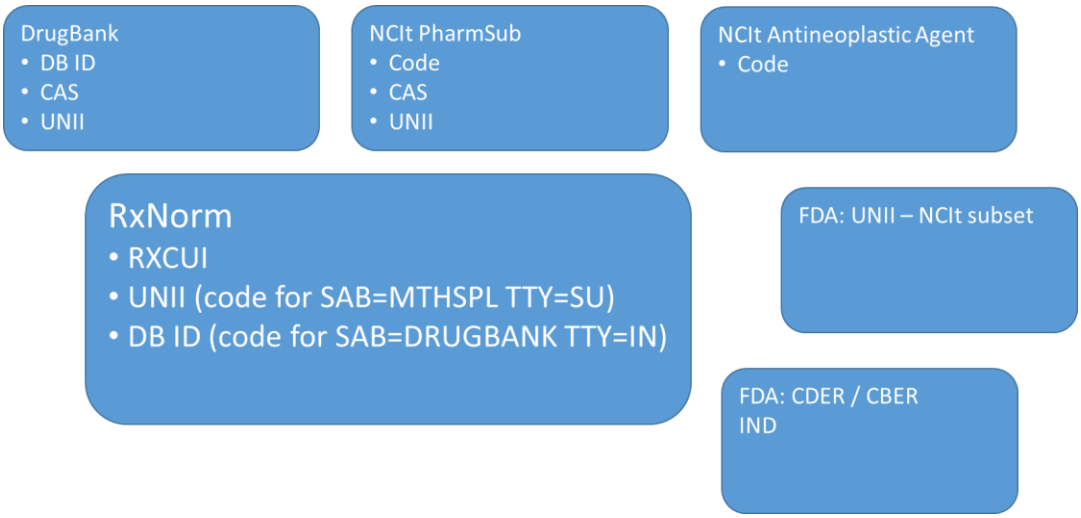

Figure 1. Sources for collecting investigational drugs and their synonyms.

Investigational drugs will be created as non-standard concepts with their synonyms in the concept_synonym table. The vocabulary for investigational drugs will be treated similarly to a non-US drug data dictionary so that a standard target concept as RxNorm Extension entry will be created if there is no

valid RxNorm match. As the sources provide content on “ingredient” level only, we decided to represent investigational drugs as ingredients. To indicate a transition from an investigational status to an approved drug a new relationship will be introduced with respective valid start and end dates to indicate the date of approval. This is particularly interesting for entries that are already approved but were investigational before within a reasonable period of time (which has to be determined).

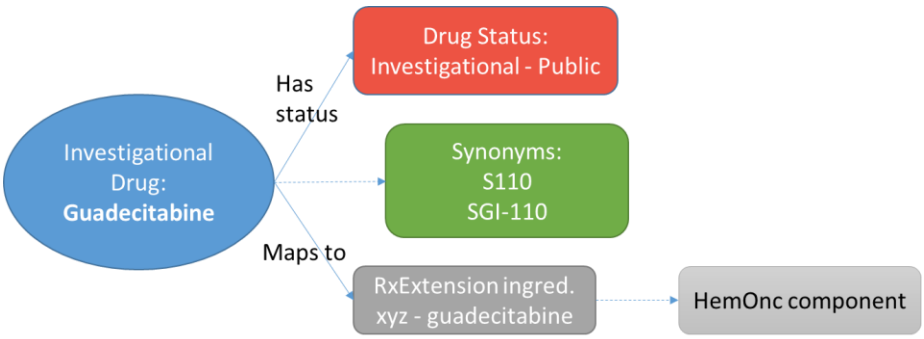

Figure 2. Model for building the investigational drug vocabulary.

This model can be extended for drugs that are private and are representing institutional intellectual property by creating local 2-billion concepts within the same ontological structure.

Results

DrugBank, as a well curated starting point, provides around 6000 entries not matched within RxNorm indicating investigational status and hence not present currently in the OMOP vocabularies. From DrugBank as well as NCIt and other sources, the respective synonyms can be collected and added to their concepts in OMOP.

Conclusion

We propose the creation of an OMOP standardized vocabulary for investigational drugs to enable researchers using existing codes in their source data for initiating network studies based not only on approved drugs. By extending this model for proprietary codes, local drugs can be included in the research as well. By introducing a drug status, previously investigational drugs can be included from historic source data based on their research designations and compared with today’s approved equivalent. However, a current status and a date of approval is not readily available from the above mentioned databases, which makes it a challenge to reliably determine this information. We hope to tap additional sources in the future to retrieve that information.

References/Citations

1. https://go.drugbank.com/
2. https://ncit.nci.nih.gov/ncitbrowser/
3. https://pubchem.ncbi.nlm.nih.gov/
4. https://chem.nlm.nih.gov/chemidplus/