**Present**

Rimma Belenkaya, Christian Reich, George Hirpcsak, Andrew Williams, Kyu-Pyo Kim, Patrick Mergler, Hamed Abedtash, Seng Chan You and two colleagues, Michael Gurley, Dmytry Dymshyts, Melanie Philofsky, Kees van Bochove, Gurvaneet Randhawa, Eric Schneider

**Discussion**

1. Presentation of G-CDM prototype – Seng Chan You
   * Model will become publically available in January
   * Group discussed where to start with genomic data; oncology is a good place
   * Need a geneticist and oncologist to join
2. Columbia NCI grant - Dr. Gurvaneet Randhawa

* Described grant goals
  + Understand the sequence of (non-cancer) treatments in cancer patients with diabetes, depression, or high blood pressure.
  + Understand the feasibility of using existing data infrastructure to conduct cancer treatment and outcomes research.
  + Detect deviations from cancer treatment guidelines
  + Analyze delivery of care
* Requested to document challenges in building OHDSI infrastructure to support cancer research

1. Discussion of current challenges, candidate approaches, and planning for the minimal viable product

* Identifying cancer diagnosis and recurrences
  + Only first occurrence is recorded in ICD-O in cancer registries
  + Some “SEER” states report all recurrences and death (what states?). Quality of reporting is better in general.
  + ICD-9/10 diagnoses in EMR may be less granular
  + ICD -9/10 diagnoses may change throughout the course of the disease
* Identifying treatment and treatment regimens
  + Individual drugs – SEER cancer drug list or RxNorm classes
  + Drug regimens – SEER regimens. We can compare with chemo orders.
  + We need cancer-specific drug eras and construction algorithms. Current algorithms will not identify treatment cycles.
  + Cancer surgeries. Detecting cancer surgeries and separating them from other surgeries. Normally, at the time when cancer treatment is delivered no other surgeries are performed. However, if we could classify cancer cases based on SNOMED treatment classes that will remove any ambiguity. Reach out to IMO for a possibility of collaboration.
  + Radiotherapies are least ambiguous
  + We can overlay cancer registry treatment schema, including modalities and order, with EMR treatment details (e.g. drugs, surgeries) to construct complete treatment regimen
  + We need to build a CDM extension to support treatment regimens/eras.
* `Staging
  + Pathology staging in a structured form is available only in cancer registry for the first occurrence
  + Clinical staging is also available in cancer registry, but is of known poor quality
  + It is not mapped to any standard terminologies
  + AJCC staging versions have ambiguous codings
  + Patrick Mergler suggested to conctact Lyn (?) from SEER

1. Identifying disease progression and response to treatemtn
   1. Imaging response to treatment (RESIST) is recorded only for solid tumors and only for cancer trials
   2. Change of treatment regimen may indicate either a recurrence or change due to toxicity
   3. Treatment adverse events is an important aspect to track
   4. Commercial NLP solutions that identify important cancer features (progression, response to treatment, etc.)
      1. FlatIron (https://flatiron.com/)
      2. ASCO CancerLinQ (<https://cancerlinq.org/>)

Andrew Stewart has done mapping oncology data into OMOP – potential presenter to the group

* + 1. Tempus (https://www.tempus.com/) – NLP pipeline, charge per patient

1. Presentation of the ICD-O to SNOMED mappings – Dmytry Dymshyts

Dmytry’s team discovered that although some of the SNOMED diagnoses corresponded verbatim to the ICD-O combo of histology and topography, their corresponding morphology and anatomy in SNOMED were less granular than ICD-O histology and topography.

Soultion

* Rely only on pre-coordination of SNOMED morphology and anatomy corresponding to ICD-O histology and topography. If a pre-coordinated SNOMED diagnosis does not exist, create OMOP-generated one.
* Present proposal for pre-coordinated terms to SNOMED.

**Next steps**

**I. Tasks to generate the minimal viable product**

1. Diagnosis
2. Vocabulary Mapping
   1. Rely only on pre-coordination of SNOMED morphology and anatomy corresponding to ICD-O histology and topography. If a pre-coordinated SNOMED diagnosis does not exist, create OMOP-generated one.
   2. Present proposal for pre-coordinated terms to SNOMED.
3. Diagnosis identification and verification
   1. Sourced from cancer registry based on ICDO-O – >SNOMED mapping
   2. Sourced from EMR based on ICD-9/10 -> SNOMED mapping or ICDO-O – >ICD-9/10 -> SNOMED
   3. Compare the diagnoses and patient identified from the two sources
   4. Explore how ICD-9/10 diagnosis changes in EMR, if at all, through the course of the disease
4. Cancer diagnosis representation in OMOP CDM – possible extension to represent recurrences
5. Staging

Choose vocabulary

1. Treatments
   1. Drugs

* Compare SEER drugs list with RxNorm cancer drug classes and decide which one should be used in the vocabulary. If SEER is more suitable, add SEER classification to the vocabulary.
* Add SEER drug cocktails to the vocabulary (only ingredients, no other specifics) in EMR based on those combinations.
* Identify chemo regimens in EMR using SEER cocktails and validate with chemo orders when available
* Work on cancer Drug Era algorithm and modeling in CDM
  1. Overlay cancer registry treatment schema (first occurrence only) with EMR details to construct complete treatment regimen
  2. Model representation of the treatment regimen (era) in the CDM

**II. Exploratory**

1. Discuss with IMO procedure mappings to SNOMED and classification of cancer cases.
2. Explore commercial solutions that claim to identify cancer treatment regimens, recurrences, progression and compare the outcomes with our attempts to identify treatment regimens
3. Invite Andrew Stewart and Mark Denise to present their efforts and results in mapping to CDM and additional modeling
4. Find a geneticist and oncologist to identify cancer-relevant genomic areas
5. Find out if we can get cancer registry data from any of the “SEER” states

**III. Operational**

1. Document challenges related to extending OHDSI infrastructure to support cancer research for NCI