How to Transform Cancer Control: Using Health Record Data in OHDSI to Facilitate Cancer Care Research

George Hripcsak, MD, MS
(Also work by RuiJun Chen, Thomas Falconer, Rimma Belenkaya, Michael Gurley)
Benefiting from Routine Care

• Complement registries
  – Co-morbidities, additional biomarkers, history
  – Broader population
  – International

• Outline
  – What is possible with electronic health records and claims databases
  – How accurate is it today for cancer care
  – What needs to be done
Observational Health Data Sciences and Informatics (OHDSI.org)

**Mission:** To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care

- >200 collaborators from 25 different countries
- Experts in informatics, statistics, epidemiology, clinical sciences
- Active participation from academia, government, industry, providers
- Over a billion records on >400 million patients in 80 databases
How OHDSI works:
Data stay local, total open science

**Source data warehouse, with identifiable patient-level data**

**Standardized, de-identified patient-level database (OMOP CDM v5)**

**Standardized large-scale analytics**

**Analysis results**

**Summary statistics results repository**

**OHDSI Data Partners**

**OHDSI Coordinating Center**

- Data network support
- Analytics development and testing
- Research and education

**OHDSI.org**
OHDSI OMOP: Deep information model with extensive vocabularies (80)
Treatment pathways for chronic disease around the world (250M population, 5 countries)

Type 2 Diabetes Mellitus
- CCAE
- CPRD
- JMDC
- Metformin
- Glitazone
- Pioglitazone
- Sitagliptin
- Glimepiride
- Glyburide
- Rosiglitazone
- Insulin, Glargine, Human
- Exenatide
- Saxagliptin

Hypertension
- CUMC
- INPC
- Losartan
- Amlodipine
- Furosemide
- Valsartan
- Carvedilol
- Triamterene
- Diltiazem
- Ramipril
- Benazepril
- Olmesartan
- Spironolactone
- Clonidine

Depression
- MDCD
- GE
- MDCR
- INPC
- OPTUM
- Citalopram
- Bupropion
- Escitalopram
- Venlafaxine
- Duloxetine
- Desvenlafaxine
- Nortriptyline
- Doxepin

Hripcsak, PNAS 2016
Improving reproducibility through large scale research

OHDSI 11% of exposure-outcome pairs have calibrated $p < 0.05$

Schuemie, Phil Trans A 2018

Literature is severely biased and 85% positive with p-value hacking

Not significant
How often do drugs have side effects

- All drugs on the world market
- All side effects
- Absolute risk in 10+ large databases

How Often...

How often do patients get a condition after starting a drug?

Which drug are you interested in?
Lisinopril

Which condition are you interested in?
Angioedema

What this does
Use this tool to look up the proportion of people starting a drug who are newly diagnosed with a condition within 1 year of starting the drug. You can search for a specific drug-condition incidence by entering your drug and condition of interest in the fields above. Or, you can browse a list of conditions of potential interest by leaving the condition field blank, and you'll be shown conditions listed on the drug’s product label.

What this does not do
This tool does not demonstrate that a drug causes a condition (i.e., that the condition is a side effect of the drug). Instead, for example, the condition may be part of the reason you are taking the drug, or the condition may just be common in the population.

This tool provides the overall observed risk in a population, but does not provide the attributable risk due to drug exposure. The results provided are raw unadjusted numbers for each diagnosis. The data made available through this site are for informational purposes only and are not a substitute for professional medical advice or services. You should not use this information for comparing drugs or making decisions related to diagnosing or treating a medical or health condition; instead, please consult a physician or healthcare professional in all matters related to your health.
NCI contract: OHDSI accuracy for cancer diagnosis and treatment

- **PPV 95.9%** from manual review
  - Manual chart review of 100 random patients
- **Sensitivity 99%** using treatment exposure
  - 1589/1599 (99.4%) patients who got cisplatin correctly identified as having cancer
- **Specificity (calculated): 99.87%**
• **Source data challenges**
  – Unlike cancer registries, EHR data are not cleaned and abstracted, but limited in time and features

• **Modeling and terminology challenges**
  – Mismatch between models, vocabularies, and conventions used in cancer versus EHR
  – Staging, histopathology, primary/secondary, recurrence, treatment regimens

• **Analytical derivation of the key disease features challenge**
  – Infer cancer recurrences and progression of disease to identify treatment episodes and response to treatment
Map between oncology (ICD-O) and general (SNOMED)

Crosswalk:
1. Concatenate ICD-O histology and topology concept IDs: 8010/3-C50.9
   You may need to make additional transformations: inserting dot, removing extra letters, etc.

2. Map the concatenated code to its OMOP concept ID: 8010/3-C50.9 (Carcinoma of Breast structure)

3. -> Maps to -> concept 4116071 (id), 254838004 (code) Carcinoma of breast

4. Use Athena to find this concept, just follow the rule that icd-o combination concept is made as <morphology>-<topography code with ".">
Extend OMOP data model cancer

- **event_occurrence_id** is a reference to an event the modifier modifies, in this case condition_occurrence_id or condition_era_id
- **event_concept_id** is a table an event is stored in, in this case ‘CONDITION_OCCURRENCE’ or ‘CONDITION_ERA’
- **condition_era_recurrence_flag** (Y/N) indicates if an era record represents cancer first occurrence or recurrence
- **condition_era_type_concept_id** identifies method of era derivation.
- **diagnostic_method_id** indicates how cancer diagnosis/diagnostic feature was diagnosed (e.g., pathology, symptomatically, record abstraction, etc.)
Proof of concept: depression treatment pathways in cancer care
To be done

• Standards for vocabulary and data model
  – Align oncology and general standards
  – Engage NLM Lister Hill Center, OHDSI

• Cancer phenotyping
  – Treatment regimens, response to treatment, remission, recurrence, progression

• Run OHDSI cancer studies

• Reduce rift between cancer informatics and general informatics