Vocabulary tutorial
Theory, principles, and practical applications

OHDSI Symposium 2016
Everything is a concept....everything needs to be defined in a common language

**Cardiovascular, Bleeding, and Mortality Risks in Elderly Medicare Patients Treated With Dabigatran or Warfarin for Nonvalvular Atrial Fibrillation**

David J. Graham, MD, MPH; Marsha E. Reichman, PhD; Michael Wernecke, BA; Rongmei Zhang, PhD; Mary Ross Southworth, PharmD; Mark Levenson, PhD; Ting-Chang Sheu, MPH; Katrina Mott, MHS; Margie R. Goulding, PhD; Monika Houstoun, PharmD, MPH; Thomas E. Macurdy, PhD; Chris Worrall, BS; Jeffrey A. Kelman, MD, MMSc

*Background*—The comparative safety of dabigatran versus warfarin for treatment of nonvalvular atrial fibrillation in general practice settings has not been established.

*Methods and Results*—We formed new-user cohorts of propensity score–matched elderly patients enrolled in Medicare who initiated dabigatran or warfarin for treatment of nonvalvular atrial fibrillation between October 2010 and December 2012. Among 134,414 patients with 37,587 person-years of follow-up, there were 2,715 primary outcome events. The hazard ratios (95% confidence intervals) comparing dabigatran with warfarin (reference) were as follows: ischemic stroke, 0.80 (0.67–0.96); intracranial hemorrhage, 0.34 (0.26–0.46); major gastrointestinal bleeding, 1.28 (1.14–1.44); acute myocardial infarction, 0.92 (0.78–1.08); and death, 0.86 (0.77–0.96). In the subgroup treated with dabigatran 75 mg twice daily, there was no difference in risk compared with warfarin for any outcome except intracranial hemorrhage, in which case dabigatran risk was reduced. Most patients treated with dabigatran 75 mg twice daily appeared not to have severe renal impairment, the intended population for this dose. In the dabigatran 150-mg twice daily subgroup, the magnitude of effect for each outcome was greater than in the combined-dose analysis.

*Conclusions*—In general practice settings, dabigatran was associated with reduced risk of ischemic stroke, intracranial hemorrhage, and death and increased risk of major gastrointestinal hemorrhage compared with warfarin in elderly patients with nonvalvular atrial fibrillation. These associations were most pronounced in patients treated with dabigatran 150 mg twice daily, whereas the association of 75 mg twice daily with study outcomes was indistinguishable from warfarin except for a lower risk of intracranial hemorrhage with dabigatran. *(Circulation. 2015;131:157–164. DOI: 10.1161/CIRCULATIONAHA.114.012061.)*

**Key Words:** anticoagulant ■ pharmacoepidemiology ■ safety ■ thrombin inhibitor ■ warfarin
OHDSI Approach

• Comprehensive
  – All of medicine and the entire world

• Don’t create yet another vocabulary
1. Select vocabularies
2. Map among vocabularies
3. Exploit existing classification hierarchies
Domains

- Condition
- Currency
- Device
- Drug
- Ethnicity
- Gender
- Measurement
- Measurement Value
- Measurement Value Operator
- Metadata
- Modifier
- Observation
- Place of Service

- Procedure
- Provider Specialty
- Race
- Relationship
- Revenue Code
- Route Of Administration
- Specimen
- Specimen Anatomic Site
- Specimen Disease Status
- Type Concept
- Unit
- Visit
- Combination Domains
OHDSI Approach

Classification hierarchies

Non-standard source terms → Standard terms → OHDSI database

Original patient data
### CONDITION_OCCURRENCE table

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>condition_concept_id</td>
<td>OMOP “201254” (SNOMED 46635009)</td>
</tr>
<tr>
<td>condition_type_concept_id</td>
<td></td>
</tr>
<tr>
<td>condition_source_value</td>
<td>ICD9CM “250.01”</td>
</tr>
<tr>
<td>condition_source_concept_id</td>
<td>OMOP “44820682” (ICD9CM 250.01)</td>
</tr>
</tbody>
</table>

**Original patient data**

- ICD9CM “250.01”
Semantic Consolidation in UMLS vs in OHDSI

A separate and independent UMLS concept represents the meaning.

One well-defined code from one vocabulary becomes the Standard Concept.
Standard terms: mapping

For every code that exists there is a **map** to a Standard Concept (including 0 if no useful mapping is possible)

- **Existing maps**
  - NDC to RxNorm
  - ICD-9-CM to SNOMED
  - SNOMED to MedDRA
  - CPT-4 to SNOMED
  - Read to SNOMED
  - ICD-9-Proc to SNOMED
  - ICD-9-Proc, CPT-4 and HCPCS to RxNorm (procedure drugs)
  - ICD-10-CM to SNOMED
  - DPD to RxNorm/Extension

- **Need**
  - OCPS-4 to SNOMED
  - Comprehensive CPT-4, LOINC, OCPS-4 and HCPCS to SNOMED

- **Working on**
  - ICD10PCS to SNOMED
  - DM+D to RxNorm/Extension
  - Gemscript to RxNorm/Extension
  - AMIS to RxNorm/Extension
  - JDBC to RxNorm/Extension
  - Other national drug schemes to RxNorm/E
  - Other national ICD-10 dialects to SNOMED
  - HCPCS to all sorts of things
  - Units to UCUM
Standard terms: one domain

For every Standard Concept exists one **Domain**
Non-standard ones can be multi-Domain

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>SNOMED Code</th>
<th>Description</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCPCS G8879</td>
<td>Clinically node negative (t1n0m0) or t2n0m0) invasive breast cancer</td>
<td>SNOMED 254837009</td>
<td>Malignant tumor of breast</td>
<td>Condition</td>
</tr>
<tr>
<td>ICD9CM V67.01</td>
<td>Following surgery, follow-up vaginal pap smear</td>
<td>SNOMED 440615002</td>
<td>Postoperative care</td>
<td>Procedure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SNOMED 133899007</td>
<td>Microscopic examination of vaginal Papanicolaou smear</td>
<td>Measurement</td>
</tr>
<tr>
<td>CPT4 90655</td>
<td>Influenza virus vaccine, split virus, preservative free, for children 6-35 months of age, for intramuscular use</td>
<td>CPT4 90655</td>
<td>Influenza virus vaccine, split virus, preservative free, for children 6-35 months of age, for intramuscular use</td>
<td>Procedure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RxNorm 5806</td>
<td>Influenza virus vaccine</td>
<td>Drug</td>
</tr>
</tbody>
</table>
Standard terms: no duplicates

For every medical entity (condition, drug, procedure etc), there is **only one Standard Concept**

- **Drug**: easy unique combination of ingredient, strength, form, and we got RxNorm, but
  - Forms are not unambiguous
  - Ingredients are easy for patented drugs, but hard for herbal, traditional, excipients, etc
  - Strength is not uniform (%, vol-%, g%, mg/dL)
  - RxNorm is US-only
- **Conditions, lab tests**: harder
  - SNOMED is trying, but
    - Duplications (4 times "Leukemic infiltration of skin")
    - Constant churn of introduction and deprecation
    - Local SNOMEDs don't help
  - LOINC good for clinical labs, too detailed for clinicians and researchers
- **Procedures, observations**: hardest
  - Procedure code systems not comprehensive, cross-links between procedures sporadic and unreliable
  - Observations: Wild West
- **Specialties, place of service**: Messy
- **Devices, disposables**: Impossible
Authoring and maintenance require the classification hierarchy

- **Infection**
  - **Lung infection**
  - **GI infect**
  - **Pneumonia**
    - **Lobar pneumonia**
  - **Pleural inf**
For every medical domain (condition, drug, procedure etc), there is a **comprehensive hierarchy**

- **Drug:** Well established and clinically used drug classes, but
  - No authority or agreement what falls under
  - Many parallel classification systems
  - Many drugs not covered
  - RxNorm has no classes

- **Conditions, Procedures, Tests:**
  - SNOMED is trying, but sometimes contorted lattice
    - Between "Neoplasm and/or hamartoma" and "Suprasellar germ cell tumor" are 3 to 11 levels of separation
  - MedDRA easy to use, but duplications and overlaps
    - "Non-site specific gastrointestinal haemorrhages", "Gastrointestinal haemorrhages"
  - CPT4: 252 codes have no hierarchical connections

- **Observations, Devices**
  - No meaningful hierarchies
Maintenance

• Long list of codes is hard to maintain
  • 312327, 319039, 434376, 436706, 438170, 438438, 438447, 441579, 444406, 4011131, 4051874, 4108669, 4119456, 4119457, 4119943, 4119944, 4119945, 4119946, 4119947, 4119948, 4121464, 4121465, 4121466, 4124684, 4124685, 4126801, 4145721, 4147223, 4151046, 4178129, 4243372, 4267568, 4270024, 4275436, 4296653, 4303359, 4324413, 43020460, 43020461, 44782712, 44782769, 45766075, 45766076, 45766115, 45766116, 45766150, 45766151, 45771322, 46270158, 46270159, 46270160, 46270161, 46270162, 46270163, 46270164, 46273495, 46274044

• Shorter list of classes that include many codes in the hierarchy
  – 312327 (SNOMED 57054005 = Acute myocardial infarction)
How well did I do?

1. Get the codes right
   – Myocardial infarction 410.00, 410.01, 410.02, ...

2. Get the cohort right
   – Patient #234, #546, #768, ...
     • “All these extra codes”
     • “Just missing one code”

3. Get the analytic result right
   – Statistical association with drug X
Vocabulary classifications improve your efficiency....and your quality

Applying standardized drug terminologies to observational healthcare databases: a case study on opioid exposure

Frank J. DeFalco · Patrick B. Ryan · M. Soledad Cepeda

- 60% of medication codes and 94% of records are mapped
- 45% of opiate codes that are covered by one of ATC, ETC, or NDF-RT are covered by all three
  - 15% missed by at least one
- No one classification scheme was better than the others
- Without classification it is hopeless
  - Consider using multiple classifications
If we try to speak the same language, will there be loss in translation?

1. Changing language may change your codelist, that may change your cohort depending on the disease

2. But in practice, running an estimation analysis using source vs. standard vocabulary yields the same result
Lessons

• Use classes to ease maintenance
  – Enumerate the classes’ codes and review

• Easier to figure out what added than what missed
  – Classes help

• Use standard terms
  – Some loss, but some gain and can be used elsewhere