OHDSI Tutorial: Design, implementation, and evaluation of cohort definitions in observational healthcare data

Patrick Ryan, PhD, Vice President of Observational Health Data Analytics at Janssen Research & Development; Adjunct Assistant Professor of Biomedical Informatics at Columbia University
Disclosures

• PBR is an employee of Janssen Research and Development, and shareholder in Johnson & Johnson

• Any opinions of the presenters expressed are their own
Agenda

1. Motivation for standardizing the cohort definition process
2. Defining a cohort in ATLAS
3. Defining a cohort in Criteria2Query
4. Hands-on experience using ATLAS and Criteria2Query
5. Evaluating cohort definitions using PheValuator
6. The journey ahead for phenotyping
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. McCurdy, 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
• Baseline characterization of target and comparator cohort

• Descriptive summaries of:
  – Demographics
  – Medical history (prior conditions)
  – Medication use (prior drugs)
  – Prior procedures
  – Risk scores
Incidence rate during target and comparator cohorts based on observing new events during ‘time-at-risk’ for eight selected outcome cohorts.
• Population-level effect estimation examining temporal association between target and comparator cohorts and eight selected outcome cohorts
The common building block of all observational analysis: cohorts

**Required inputs:**

- **Target cohort:** Person
  - cohort start date
  - cohort end date

- **Comparator cohort:** Person
  - cohort start date
  - cohort end date

- **Outcome cohort:** Person
  - cohort start date
  - cohort end date

**Desired outputs:**

- **Clinical characterization**
  - Baseline summary of exposures (treatment utilization)

- **Clinical characterization**
  - Baseline summary of outcome (disease natural history)

- **Incidence summary**
  - Proportion/rate of outcome occurring during time-at-risk for exposure

- **Population-level effect estimation**
  - Relative risk (HR, OR, IRR) of outcome occurring during time-at-risk for exposure

- **Patient-level prediction**
  - Probability of outcome occurring during time-at-risk for each patient in population
## Graham replication: Cohort characterization in ATLAS

### Features are baseline characteristics (e.g., collected before/on cohort start)

**Long Term:** 365 day lookback. **Short Term:** 30d lookback. **Overlapping:** Event spans cohort start date.

<table>
<thead>
<tr>
<th>Concept Name</th>
<th>Time Window</th>
<th>Person Count</th>
<th>% of cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>dabigatan etexilate</td>
<td>Long Term</td>
<td>19,975</td>
<td>100.00</td>
</tr>
<tr>
<td>Metoprolol</td>
<td>Long Term</td>
<td>8,820</td>
<td>44.20</td>
</tr>
<tr>
<td>Hydrochlorothiazide</td>
<td>Long Term</td>
<td>5,955</td>
<td>29.90</td>
</tr>
<tr>
<td>Acetaminophen</td>
<td>Long Term</td>
<td>5,739</td>
<td>28.80</td>
</tr>
<tr>
<td>Lisinopril</td>
<td>Long Term</td>
<td>4,935</td>
<td>24.80</td>
</tr>
<tr>
<td>Simvastatin</td>
<td>Long Term</td>
<td>4,851</td>
<td>24.30</td>
</tr>
<tr>
<td>Amlodipine</td>
<td>Long Term</td>
<td>4,808</td>
<td>24.10</td>
</tr>
<tr>
<td>Furosemide</td>
<td>Long Term</td>
<td>4,795</td>
<td>24.10</td>
</tr>
<tr>
<td>Hydrocodone</td>
<td>Long Term</td>
<td>4,590</td>
<td>23.00</td>
</tr>
<tr>
<td>atorvastatin</td>
<td>Long Term</td>
<td>4,422</td>
<td>22.20</td>
</tr>
</tbody>
</table>
Graham replication: Incidence summary design in ATLAS
Graham replication:
Incidence summary implementation in ATLAS

<table>
<thead>
<tr>
<th>Source</th>
<th>Name</th>
<th>Persons</th>
<th>Cases</th>
<th>Proportion [+/-] per 1k persons</th>
<th>Time At Risk (years)</th>
<th>Rate [+/-] per 1k years</th>
<th>Started</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUENMDCR_V657</td>
<td>Truven MDCR</td>
<td>18,376</td>
<td>93</td>
<td>5.06</td>
<td>5.852</td>
<td>15.89</td>
<td>2017-12-02, 22:46</td>
<td>00:00:29</td>
</tr>
</tbody>
</table>


Summary Statistics:

<table>
<thead>
<tr>
<th>Persons</th>
<th>Cases</th>
<th>Proportion [+/-] per 1k persons</th>
<th>Time At Risk (years)</th>
<th>Rate [+/-] per 1k years</th>
</tr>
</thead>
<tbody>
<tr>
<td>18,376</td>
<td>93</td>
<td>5.06</td>
<td>5.852</td>
<td>15.89</td>
</tr>
</tbody>
</table>

Stratify Rule:

<table>
<thead>
<tr>
<th>Rule</th>
<th>N</th>
<th>Cases</th>
<th>Proportion [+/-] per 1k persons</th>
<th>Time At Risk (years)</th>
<th>Rate [+/-] per 1k years</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender = MALE</td>
<td>10,453</td>
<td>50</td>
<td>4.78</td>
<td>3.391</td>
<td>14.74</td>
</tr>
<tr>
<td>age &lt; 75</td>
<td>7,897</td>
<td>27</td>
<td>3.42</td>
<td>2.508</td>
<td>10.77</td>
</tr>
</tbody>
</table>

34 cases, 1,580 TAR, Rate: 21.52 per 1k years 5,097 (27.74%) people. 0 criteria passed, 2 criteria failed.
Graham replication: Population-level effect estimation design in ATLAS
Graham replication:
Population-level effect estimation implementation using OHDSI methods

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>lower .95</th>
<th>upper .95</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>0.89626</td>
<td>0.71863</td>
<td>1.11829</td>
</tr>
</tbody>
</table>

Population counts
<table>
<thead>
<tr>
<th></th>
<th>treatedPersons</th>
<th>comparator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>17460</td>
<td>17460</td>
</tr>
</tbody>
</table>

Outcome counts
<table>
<thead>
<tr>
<th></th>
<th>treatedPersons</th>
<th>comparator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>164</td>
<td>155</td>
</tr>
</tbody>
</table>

Time at risk
<table>
<thead>
<tr>
<th></th>
<th>treatedDays</th>
<th>comparator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days</td>
<td>4912947</td>
<td>3954046</td>
</tr>
</tbody>
</table>

Kaplan-Meier Plot

Survival probability vs. Time in days

Legend:
- Dabigatran
- Warfarin
Defining cohorts
Defining ‘phenotype’

A phenotype is a specification of an observable, potentially changing state of an organism (as distinguished from the genotype, derived from genetic makeup).

The term phenotype can be applied to patient characteristics inferred from electronic health record (EHR) data.

The goal is to draw conclusions about a target concept based on raw EHR data, claims data, or other clinically relevant data.

Phenotype algorithms – ie, algorithms that identify or characterize phenotypes – may be generated by domain exerts and knowledge engineers, or through diverse forms of machine learning to generate novel representations of data.
Two Approaches to Phenotyping

- Rule-Based Phenotyping
- Probabilistic Phenotyping
Data are Like Lego Bricks for Phenotyping

- Conditions
- Drugs
- Procedures
- Measurements
- Observations
- Visits
Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance

Wei-Qi Wei¹, Pedro L Teixeira¹, Huan Mo¹, Robert M Cronin¹,², Jeremy L Warner¹,², Joshua C Denny¹,²

Figure 1: Weighted Venn diagrams of the distributions of patients with ICD-9, primary notes, and specific medications. Each color represents a resource. Different area colors represent the number of patients that were found within intersecting resources.
Database queries for hospitalizations for acute congestive heart failure: flexible methods and validation based on set theory

Marc Rosenman,¹,² Jinghua He,³ Joel Martin,² Kavitha Nutakki,¹ George Eckert,⁴ Kathleen Lane,⁴ Irmina Gradus-Pizlo,⁵ Siu L Hui²,⁴

**Table 3** Results for the 10 congestive heart failure (CHF) phenotype queries

<table>
<thead>
<tr>
<th>Criteria to combine Venn diagram zones</th>
<th>N in query</th>
<th>Sensitivity (%)</th>
<th>Sensitivity, SE (%)</th>
<th>PPV (%)</th>
<th>PPV, SE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any CHF</td>
<td>66 942</td>
<td>94.3</td>
<td>1.3</td>
<td>42.8</td>
<td>1.5</td>
</tr>
<tr>
<td>Any dx of 428</td>
<td>64 832</td>
<td>90.9</td>
<td>1.3</td>
<td>42.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Any dx of CHF and BNP &gt;500 pg/mL</td>
<td>21 801</td>
<td>50.8</td>
<td>1.8</td>
<td>70.7</td>
<td>2.5</td>
</tr>
<tr>
<td>¹º dx of any CHF</td>
<td>19 339</td>
<td>54.8</td>
<td>1.9</td>
<td>86.0</td>
<td>2.2</td>
</tr>
<tr>
<td>¹º dx of 428</td>
<td>16 724</td>
<td>47.6</td>
<td>1.7</td>
<td>86.3</td>
<td>2.5</td>
</tr>
<tr>
<td>¹º dx of any CHF and BNP &gt;500 pg/mL</td>
<td>11 298</td>
<td>33.5</td>
<td>1.3</td>
<td>90.0</td>
<td>2.1</td>
</tr>
<tr>
<td>¹º dx of 428 and BNP &gt;500 pg/mL</td>
<td>9662</td>
<td>28.8</td>
<td>1.1</td>
<td>90.4</td>
<td>2.4</td>
</tr>
<tr>
<td>¹º dx of 428 and BNP &gt;500 pg/mL and echocardiogram</td>
<td>5678</td>
<td>16.2</td>
<td>0.8</td>
<td>86.6</td>
<td>3.5</td>
</tr>
<tr>
<td>¹º dx of any CHF or BNP &gt;500 pg/mL</td>
<td>29 587</td>
<td>71.4</td>
<td>2.1</td>
<td>73.3</td>
<td>2.2</td>
</tr>
<tr>
<td>¹º dx of 428 or BNP &gt;500 pg/mL</td>
<td>28 863</td>
<td>69.6</td>
<td>2.1</td>
<td>73.2</td>
<td>2.2</td>
</tr>
<tr>
<td>High BNP, no ICD-9 diagnosis for CHF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zone X: no ICD-9 dx of 428, but BNP &gt;500 pg/mL</td>
<td>12 149</td>
<td>N/A</td>
<td>N/A</td>
<td>14.3</td>
<td>3.5</td>
</tr>
</tbody>
</table>

BNP, B-natriuretic peptide; PPV, positive predictive value.
OHDSI’s definition of ‘cohort’

Cohort = a set of persons who satisfy one or more inclusion criteria for a duration of time

Objective consequences based on this cohort definition:
• One person may belong to multiple cohorts
• One person may belong to the same cohort at multiple different time periods
• One person may not belong to the same cohort multiple times during the same period of time
• One cohort may have zero or more members
• A codeset is NOT a cohort...
  ...logic for how to use the codeset in a criteria is required
Dissecting the anatomy of a cohort definition

Person timeline

Cohort entry
Initial event

Cohort exit

Observation period end

Observation period start

Inclusion criteria

- Observation (>=1)
- Absence (=0)

Temporal logic
Questions to answer when defining a cohort

- What initial event(s) define cohort entry?
- What inclusion criteria are applied to the initial events?
- What defines a person’s cohort exit?
What initial event(s) define cohort entry?

- Events are recorded time-stamped observations for the persons, such as drug exposures, conditions, procedures, measurements and visits.
- The event index date is set to be equal to the event start date.
- Initial events defined by a domain, conceptset, and any domain-specific attributes required.
What inclusion criteria are applied to the initial events?

- The qualifying cohort will be defined as all persons who have an initial event and satisfy all qualifying inclusion criteria.
- Each inclusion criteria is defined by domain(s), conceptset(s), domain-specific attributes, and the temporal logic relative to initial events.
- Each qualifying inclusion criteria can be evaluated to determine the impact of the criteria on the attrition of persons from the initial cohort (example use case: clinical trial feasibility).
What defines a person’s cohort exit?

- Cohort exit signifies when a person no longer qualifies for cohort membership
- Cohort exit can be defined in multiple ways:
  - End of observation period
  - Fixed time interval relative to initial event
  - Last event in a sequence of related observations (ex: persistent drug exposure)
  - Censoring observations
- Cohort exit strategy will impact whether a person can belong to the cohort multiple times during different time intervals
Defining cohort components

- **Domain**: A Domain defines the set of allowable Concepts for the standardized fields in the CDM tables.
  - Ex: Condition, Drug, Procedure, Measurement

- **Conceptset**: An expression that defines one or more concepts encompassing a clinical entity of interest
  - Ex: Concepts for T2DM, concepts for antidiabetic drugs

- **Domain-specific attribute**:  
  - Ex: DRUG_EXPOSURE: Days supply; MEASUREMENT: value_as_number, high_range

- **Temporal logic**: the time intervals within which the relationship between an inclusion criteria and an event is evaluated
  - Ex: Indicated condition must occur during 365d prior to or on exposure start
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 134414 patients with 37587 person-years of follow-up, there were 2715 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
Graham et al. description of the outcomes

**Study Outcomes**

The primary outcomes were ischemic stroke, major bleeding with specific focus on intracranial and gastrointestinal bleeding, and AMI. Secondary outcomes were all hospitalized bleeding events and mortality. The *International Classification of Diseases, Ninth Revision, Clinical Modification* codes used to define these outcomes are listed in Table II in the online-only Data Supplement. The codes defining ischemic stroke have a positive predictive value (PPV) of 88% to 95%.\(^{18-20}\) Major bleeding was defined as

![Table 2](image-url)

*Table 2. International Classification of Disease, 9th edition. Clinical Modification (ICD 9-CM) codes used to define study outcomes.*
Exercise: Define the outcome cohort for Graham et al.

• What initial event(s) define cohort entry?
• What inclusion criteria are applied to the initial events?
• What defines a person’s cohort exit?
A new-user retrospective cohort design was used to compare patients initiating dabigatran or warfarin for the treatment of nonvalvular AF. We identified all patients with any inpatient or outpatient diagnoses of AF or atrial flutter based on International Classification of Diseases, Ninth Revision coding who also filled at least 1 prescription for either drug from October 19, 2010 (US dabigatran approval date) through December 31, 2012, the study end date. Patients were excluded if they had <6 months of enrollment in Medicare before their index dispensing, were aged <65 years, received prior treatment with a study medication or rivaroxaban or apixaban (anticoagulants approved during the study), were in a skilled nursing facility or nursing home, or were receiving hospice care on the date of their cohort-qualifying prescription. Patients were also excluded if they had a hospitalization that extended beyond the index dispensing date. Patients discharged from the hospital on the same day as their index dispensing were included. Patients undergoing dialysis and kidney transplant recipients were also excluded. Additionally, because warfarin is approved for indications other than AF, we excluded patients with diagnoses indicating the presence of mitral valve disease, heart valve repair or replacement, deep vein thrombosis, pulmonary embolism, or joint replacement surgery in the preceding 6 months.
Exercise: Define the target exposure cohort for Graham et al.

- What initial event(s) define cohort entry?
- What inclusion criteria are applied to the initial events?
- What defines a person’s cohort exit?
What initial event(s) define cohort entry?

• Do:
  – Define by existence of any observation in any domain

• Don’t:
  – Define by absence of an observation - when does absence occur?
  – Define by age- year of birth is constant, but requires index date to anchor age calculation

• Caution:
  – Defining a cohort by calendar date can cause observation bias, since that date unlikely to be at point of health service utilization, ex: cases matched to controls. Consider instead defining by a visit that occurs within a calendar timeframe.
What inclusion criteria are applied to the initial events?

- **Do:**
  - Specify all criteria as inclusion criteria to avoid confusion of Boolean logic around inclusion vs. exclusion
  - Use information on or before index event (think like a randomized trial: index event is study start, can’t predict future)

- **Don’t:**
  - Assume temporal logic, but always provide relative time window to evaluate criteria

- **Caution:**
  - There’s a difference between ‘first time in history with >365d prior observation’ vs. ‘no prior observation in last 365 days’
  - One person may have multiple initial events, criteria are applied to each event (not person)
What defines a person’s cohort exit?

• Do:
  – Specify a cohort exit, even if you are not intending to use it for your analytic use case

• Don’t:
  – Confuse censoring for analytical purposes with cohort definition (which can be analysis-independent)…ex: censoring at time of outcome

• Caution:
  – Time-of-cohort participation can be different from analysis time-at-risk…ex: acute effects can be studied using a fixed window post-exposure start, intent-to-treat analysis can follow person through observation period end
Defining a cohort in ATLAS

Chris Knoll
Janssen Research and Development
Defining a cohort using Criteria2Query

Cong Liu
Columbia University Medical Center
Evaluating a phenotype using PheValuator

Patrick Ryan
Janssen Research and Development
Columbia University Medical Center
Questions?

Thanks for joining the journey!

ryan@ohdsi.org