



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 PheEvaluator
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. 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



- Baseline characterization of target and comparator cohort
- Descriptive summaries of:
 - Demographics
 - Medical history (prior conditions)
 - Medication use (prior drugs)
 - Prior procedures
 - Risk scores

Table 1. Sociodemographic Factors, Medical Conditions, and Medication Use at Baseline in Propensity Score-Matched Medicare Beneficiaries Initiating Dabigatran or Warfarin for Atrial Fibrillation, 2010–2012

| Characteristic | Dabigatran, % (n=67 207) | Warfarin, % (n=67 207) | Standardized Mean Difference |
|-------------------------------|-----------------------------|---------------------------|------------------------------------|
| Age group, y | | | |
| 65–74 | 42 | 41 | 0.01 |
| 75–84 | 43 | 43 | 0.01 |
| ≥85 | 16 | 16 | 0.00 |
| Female sex | 51 | 52 | 0.01 |
| Race/ethnicity | | | |
| White | 92 | 92 | 0.00 |
| Black | 3 | 3 | 0.00 |
| Other | 5 | 5 | 0.00 |
| Medical history | | | |
| General | | | |
| Diabetes mellitus | 33 | 34 | 0.00 |
| Hypercholesterolemia | 74 | 74 | 0.00 |
| Hypertension | 87 | 87 | 0.00 |
| Kidney failure | | | |
| Acute | 5 | 5 | 0.00 |
| Chronic | 13 | 13 | 0.00 |
| Obesity | 11 | 11 | 0.00 |
| Peptic ulcer disease | <1 | <1 | 0.00 |
| Prior bleeding event | | | |
| Hospitalized | 1 | 1 | 0.00 |
| Not hospitalized | 3 | 3 | 0.01 |
| Smoking | 16 | 16 | 0.01 |
| Cardiovascular disease | | | |
| Acute myocardial infarction | | | |
| Past 1–30 d | 1 | 1 | 0.01 |
| Past 31–183 d | 1 | 1 | 0.00 |
| Coronary revascularization | 16 | 16 | 0.01 |
| Heart failure | | | |
| Hospitalized | 4 | 4 | 0.01 |
| Outpatient | 14 | 14 | 0.00 |
| Other ischemic heart disease | 48 | 49 | 0.01 |
| Stroke | | | |
| Past 1–30 d | 2 | 2 | 0.00 |
| Past 31–183 d | 1 | 2 | 0.00 |
| Other cerebrovascular disease | 13 | 13 | 0.00 |
| Transient ischemic attack | 7 | 7 | 0.00 |
| Cardioablation | 2 | 2 | 0.00 |
| Cardioversion | 9 | 9 | 0.02 |
| Other medical conditions | | | |
| Falls | 5 | 5 | 0.00 |
| Fractures | 2 | 2 | 0.00 |
| Syncope | 10 | 10 | 0.00 |
| Walker use | 3 | 3 | 0.00 |
| CHADS ₂ score* | | | |
| 0–1 | 28 | 28 | 0.01 |

Table 1. Continued

| Characteristic | Dabigatran, % (n=67 207) | Warfarin, % (n=67 207) | Standardized Mean Difference |
|-----------------------------|-----------------------------|---------------------------|------------------------------------|
| 2 | 40 | 40 | 0.00 |
| 3 | 21 | 21 | 0.01 |
| ≥4 | 10 | 11 | 0.01 |
| HAS-BLED score† | | | |
| 1 | 9 | 9 | 0.01 |
| 2 | 50 | 50 | 0.01 |
| 3 | 32 | 32 | 0.01 |
| ≥4 | 9 | 9 | 0.00 |
| Medication use | | | |
| General | | | |
| Estrogen replacement | 2 | 3 | 0.00 |
| H2 antagonists | 5 | 5 | 0.00 |
| NSAIDs | 15 | 15 | 0.00 |
| Proton pump inhibitors | 26 | 27 | 0.01 |
| SSRI antidepressants | 13 | 13 | 0.01 |
| Cardiovascular | | | |
| ACE/ARB | 59 | 59 | 0.00 |
| Antiarrhythmics | 25 | 25 | 0.01 |
| Anticoagulants (injectable) | 7 | 7 | 0.01 |
| Antiplatelets | 17 | 17 | 0.01 |
| β-Blockers | 70 | 71 | 0.00 |
| Calcium channel blockers | 42 | 42 | 0.01 |
| Digoxin | 17 | 16 | 0.00 |
| Diuretics | | | |
| Loop | 28 | 28 | 0.00 |
| Potassium sparing | 5 | 5 | 0.01 |
| Thiazide | 29 | 29 | 0.00 |
| Nitrates | 10 | 11 | 0.01 |
| Statins | 57 | 57 | 0.00 |
| Fibrates | 5 | 5 | 0.00 |
| Diabetes related | | | |
| Insulin | 6 | 6 | 0.00 |
| Metformin | 13 | 14 | 0.00 |
| Sulfonylureas | 9 | 10 | 0.00 |
| Other | 6 | 6 | 0.00 |
| Metabolic inhibitors‡ | | | |
| Amiodarone | 10 | 10 | 0.00 |
| Dronedarone | 5 | 5 | 0.02 |
| Verapamil | 2 | 2 | 0.00 |
| Azole antifungals | <1 | <1 | 0.00 |

Additional factors included in the propensity score model are shown in the online-only Data Supplement. ACE/ARB indicates angiotensin converting-enzyme inhibitor/angiotensin receptor blocker; NSAIDs, nonsteroidal anti-inflammatory drugs; and SSRI, selective serotonin reuptake inhibitor.

*The CHADS₂ score assigns points for the presence of congestive heart failure, hypertension, age ≥75 y, diabetes mellitus, stroke, or transient ischemic attack.¹¹

†The HAS-BLED score assigns points for the presence of hypertension, abnormal renal or liver function, stroke, bleeding history, labile international normalized ratio, age ≥65 y, and antiplatelet drug or alcohol use.^{12,13} Labile international normalized ratio could not be determined from claims data and was excluded from our scoring.

‡Days supply of use overlapped with the date of first prescription for warfarin.

Table 2. Outcome Event Counts, Incidence Rates, and Adjusted Hazard Ratios With 95% CIs Comparing Propensity Score-Matched New-User Cohorts of Dabigatran and Warfarin Treated for Nonvalvular Atrial Fibrillation, With Warfarin as the Reference Group

| | No. of Events | | Incidence Rate per 1000 Person-Years | | 95% CI (Dabigatran vs Warfarin) |
|-----------------------------|---------------|----------|--------------------------------------|----------|---------------------------------|
| | Dabigatran | Warfarin | Dabigatran | Warfarin | |
| Primary outcomes | | | | | |
| Ischemic stroke | 205 | 270 | 11.3 | 13.9 | |
| Major hemorrhage | 777 | 851 | 42.7 | 43.9 | |
| Gastrointestinal | 623 | 513 | 34.2 | 26.5 | |
| Intracranial | 60 | 186 | 3.3 | 9.6 | |
| Intracerebral | 44 | 142 | 2.4 | 7.3 | |
| Acute myocardial infarction | 285 | 327 | 15.7 | 16.9 | |
| Secondary outcomes | | | | | |
| All hospitalized bleeds | 1079 | 1139 | 59.3 | 58.8 | |
| Mortality* | 603 | 744 | 32.6 | 37.8 | |

*For 1064 deaths not preceded by a primary study outcome, the adjusted hazard ratio (95% confidence interval [CI]) was 0.89 (0.79–1.00; $P=0.051$), whereas for 283 deaths occurring within 30 days after a primary outcome, the adjusted hazard ratio (95% CI) was 0.77 (0.61–0.98; $P=0.03$).

- Incidence rate during target and comparator cohorts based on observing new events during ‘time-at-risk’ for eight selected outcome cohorts

Table 2. Outcome Event Counts, Incidence Rates, and Adjusted Hazard Ratios With 95% CIs Comparing Propensity Score-Matched New-User Cohorts of Dabigatran and Warfarin Treated for Nonvalvular Atrial Fibrillation, With Warfarin as the Reference Group

| | | Adjusted Hazard Ratio (95% CI) | P Value |
|-----------------------------|--|-----------------------------------|---------|
| Primary outcomes | | | |
| Ischemic stroke | | 0.80 (0.67–0.96) | 0.02 |
| Major hemorrhage | | 0.97 (0.88–1.07) | 0.50 |
| Gastrointestinal | | 1.28 (1.14–1.44) | <0.001 |
| Intracranial | | 0.34 (0.26–0.46) | <0.001 |
| Intracerebral | | 0.33 (0.24–0.47) | <0.001 |
| Acute myocardial infarction | | 0.92 (0.78–1.08) | 0.29 |
| Secondary outcomes | | | |
| All hospitalized bleeds | | 1.00 (0.92–1.09) | 0.97 |
| Mortality* | | 0.86 (0.77–0.96) | 0.006 |

*For 1064 deaths not preceded by a primary study outcome, the adjusted hazard ratio (95% confidence interval [CI]) was 0.89 (0.79–1.00; $P=0.051$), whereas for 283 deaths occurring within 30 days after a primary outcome, the adjusted hazard ratio (95% CI) was 0.77 (0.61–0.98; $P=0.03$).

- 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:

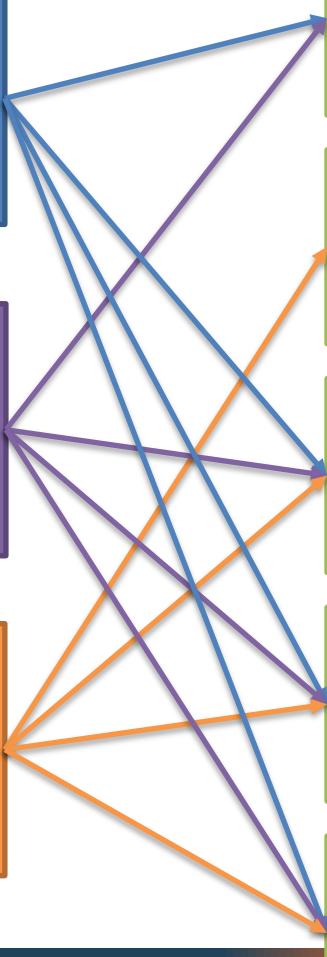
Clinical characterization
Baseline summary of exposures
(treatment utilization)

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)

Demographics Conditions Drugs Procedures Measurements Observations Distributions

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

Column visibility Copy CSV Show 15 ▾ entries

Filter:

Showing 1 to 15 of 305 entries

Previous [1](#) [2](#) [3](#) [4](#) [5](#) ... [21](#) [Next](#)

| Analysis | Concept Name | Time Window | Person Count | % of cohort |
|-------------------|----------------------|-------------|--------------|-------------|
| Group Era (1025) | | | | |
| Era (681) | dabigatran etexilate | Long Term | 19,975 | 100.00 |
| Time Window | | | | |
| Long Term (708) | Metoprolol | Long Term | 8,820 | 44.20 |
| Short Term (537) | Hydrochlorothiazide | Long Term | 5,955 | 29.90 |
| Overlapping (461) | Acetaminophen | Long Term | 5,739 | 28.80 |
| | Lisinopril | Long Term | 4,935 | 24.80 |
| | Simvastatin | Long Term | 4,851 | 24.30 |
| | Amlodipine | Long Term | 4,808 | 24.10 |
| | Eurosemide | Long Term | 4,795 | 24.10 |
| | Hydrocodone | Long Term | 4,590 | 23.00 |
| | atorvastatin | Long Term | 4,422 | 22.20 |



Graham replication: Incidence summary design in ATLAS

ATLAS

Home Data Sources Vocabulary Concept Sets Cohorts Incidence Rates Profiles Estimation Prediction Jobs Configuration Feedback

Incidence Rate Analysis

OHDSI cohort tutorial: Graham replication

Save Close Copy Delete Generate... ▾

Definition Concept Sets Generation Utilities

Study Cohorts

| Target Cohorts | Outcome Cohorts |
|---|---|
| ✗ #2649: OHDSI estimation tutorial: Graham replication: target cohort - dabigatran new users with prior atrial fibrillation | ✗ #5160: OHDSI cohort tutorial: Graham replication: outcome cohort #1 - incident ischemic stroke, observed in inpatient setting |
| ✗ #5159: OHDSI cohort tutorial: Graham replication: comparator cohort - warfarin new users with prior atrial fibrillation | ✗ #5161: OHDSI cohort tutorial: Graham replication: outcome cohort #2 - incident intracranial hemorrhage, observed in inpatient setting |
| | ✗ #5162: OHDSI cohort tutorial: Graham replication: outcome cohort #3 - incident major gastrointestinal (GI) bleeding events, observed in inpatient setting |

Add Target Cohort Add Outcome Cohort

Time At Risk

Time at risk defines the time window relative to the cohort start or end date with an offset to consider the person 'at risk' of the outcome.

- Time at risk starts with plus days.
- Time at risk ends with plus days.



Graham replication: Incidence summary implementation in ATLAS

ATLAS

Source Name Persons Cases Proportion [+|-] per 1k persons Time At Risk (years) Rate [+|-] per 1k years Started Duration

TRUVENMDCR_V657 Truven MDCR Execute 18,376 93 5.06 5,852 15.89 2017-12-02, 22:46 00:00:29 Remove

Showing target cohort: OHDSI estimation tutorial: Graham replicat ▾ and outcome cohort: OHDSI cohort tutorial: Graham replication: ▾

| Stratify Rule | N | Cases | Proportion [+ -] per 1k persons | Time At Risk (years) | Rate [+ -] per 1k years |
|------------------|--------|-------|---------------------------------|----------------------|-------------------------|
| 1. gender = MALE | 10,453 | 50 | 4.78 | 3,391 | 14.74 |
| 2. age < 75 | 7,897 | 27 | 3.42 | 2,508 | 10.77 |

Summary Statistics: 18,376 93 5.06 5,852 15.89

34 cases, 1,580 TAR, Rate: 21.52 per 1k years
5,097 (27.74%) people, 0 criteria passed, 2 criteria failed.

10.20 11.46 12.72 13.98 15.23 16.49 17.75 19.00 20.26 21.52



Graham replication: Population-level effect estimation design in ATLAS

ATLAS

Population Level Effect Estimation

OHDSI estimation tutorial: Graham replication: dabigatran vs warfarin for risk of ischemic stroke

Save Close Delete

Specification Utilities

Choose your target cohort:

OHDSI estimation tutorial: Graham replication: target cohort - dabigatran new users with prior atrial fibrillation

Choose your comparator cohort:

OHDSI estimation tutorial: Graham replication: comparator cohort - warfarin new users with prior atrial fibrillation

Choose your outcome cohort:

OHDSI estimation tutorial: Graham replication: outcome cohort #1 - incident ischemic stroke, observed in inpatient setting

Specify the statistical model used to estimate the risk of outcome between target and comparator cohorts:

Cox proportional hazards ▾

Define the time-at-risk window start, relative to target/comparator cohort entry:

1 ▾ days from cohort start date

Define the time-at-risk window end:

0 ▾ days from cohort end date ▾

Minimum washout period applied to target and comparator cohorts:

0 ▾

Minimum required days at risk, applied to target and comparator cohorts:

1 ▾

Remove patients who enter both cohorts?

No ▾

Remove patients who have observed the outcome prior to cohort entry?

Yes ▾



Graham replication: Population-level effect estimation implementation using OHDSI methods

Model type: cox

Stratified: FALSE

Use covariates: FALSE

Status: OK

| | Estimate | lower .95 | upper .95 |
|-----------|----------|-----------|-----------|
| treatment | 0.89626 | 0.71863 | 1.11829 |

Population counts

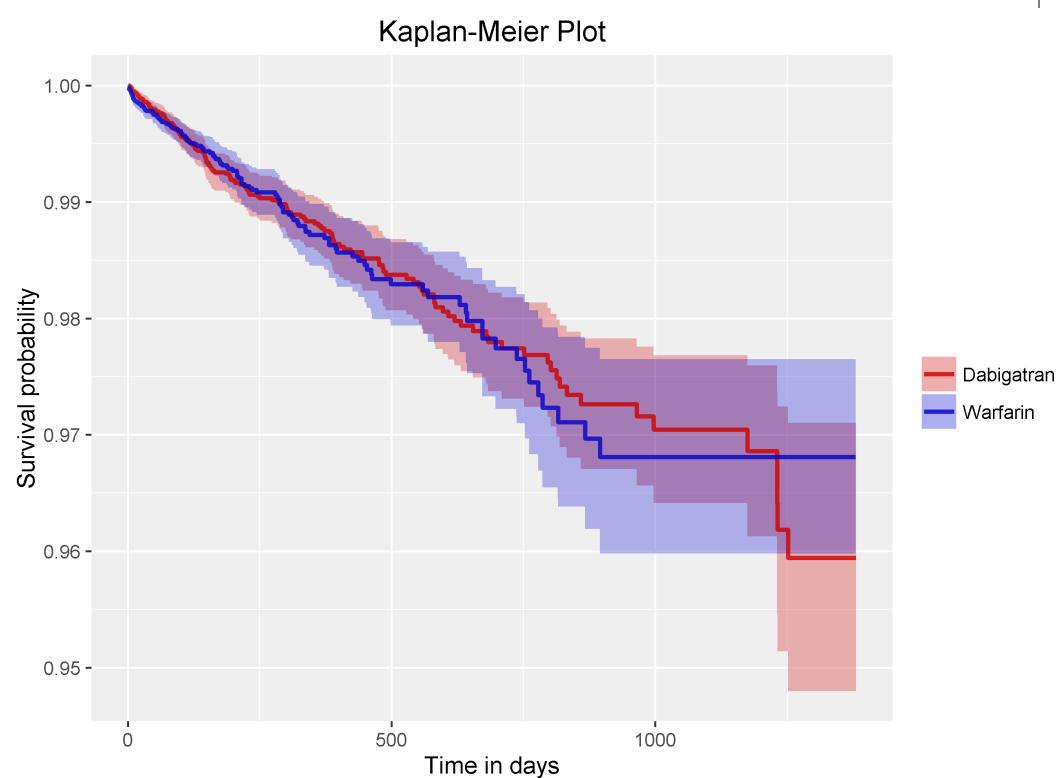
| | treatedPersons |
|-------|----------------|
| Count | 17460 |

Outcome counts

| | treatedPersons |
|-------|----------------|
| Count | 164 |

Time at risk

| | treatedDays | comparatorDays |
|------|-------------|----------------|
| Days | 4912947 | 3954046 |





Defining cohorts



Defining ‘phenotype’

Journal of the American Medical Informatics Association, 0(0), 2017, 1–6

doi: 10.1093/jamia/ocx110

Perspective



INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD

Perspective

High-fidelity phenotyping: richness and freedom from bias

George Hripcsak¹ and David J Albers¹

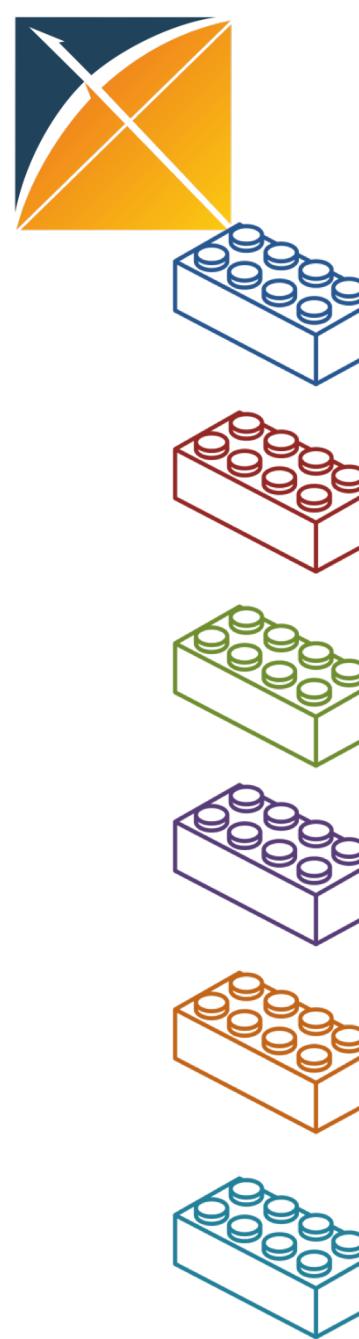
- A phenotype is a specification of an observable, potentially changing state of an organism (as distinguished from the genotype, derived from genetic makeup).
- The term phenotype can be applied to patient characteristics inferred from electronic health record (EHR) data.
- The goal is to draw conclusions 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 experts 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

RECEIVED 8 January 2015

REVISED 14 July 2015

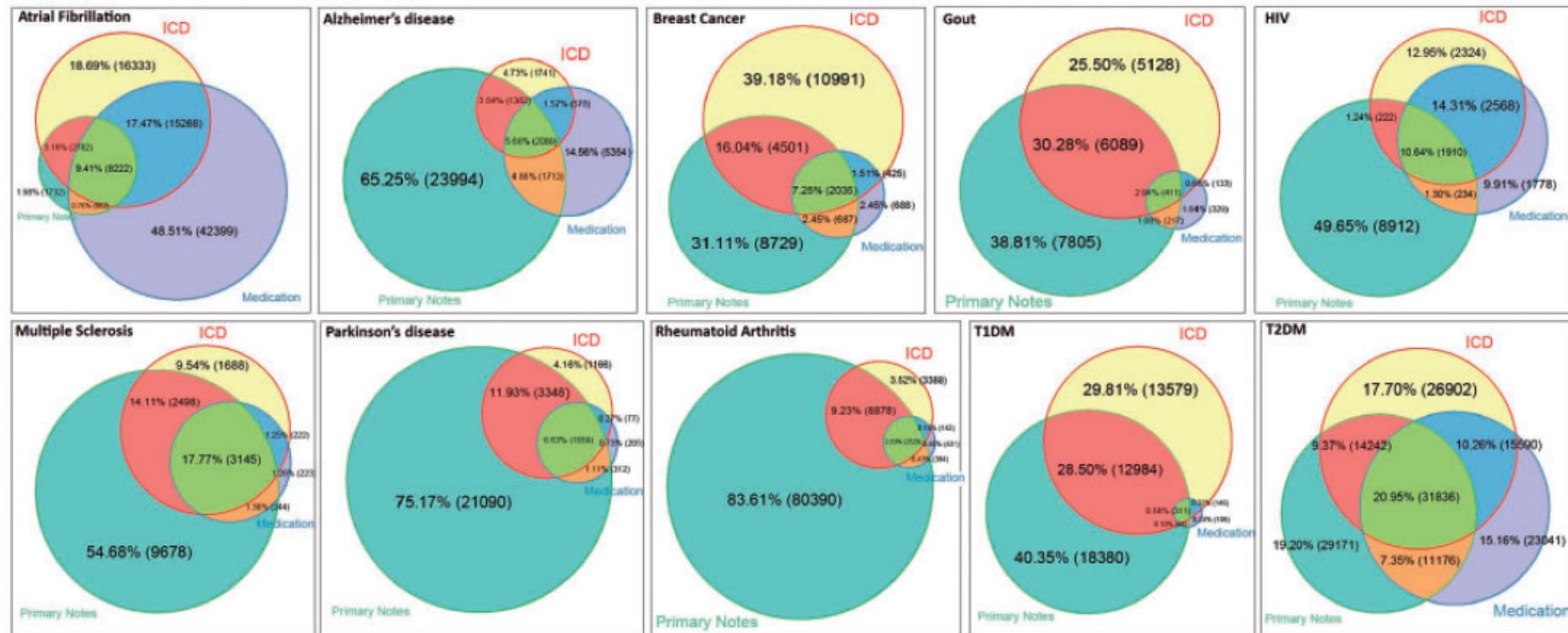
ACCEPTED 15 July 2015

PUBLISHED ONLINE FIRST 2 September 2015



Wei-Qi Wei¹, Pedro L Teixeira¹, Huan Mo¹, Robert M Cronin^{1,2}, Jeremy L Warner^{1,2}, Joshua C Denny^{1,2}

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 Hui^{2,4}



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

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

BNP, B-natriuretic peptide; PPV, positive predictive value.



OHDSI's definition of 'cohort'

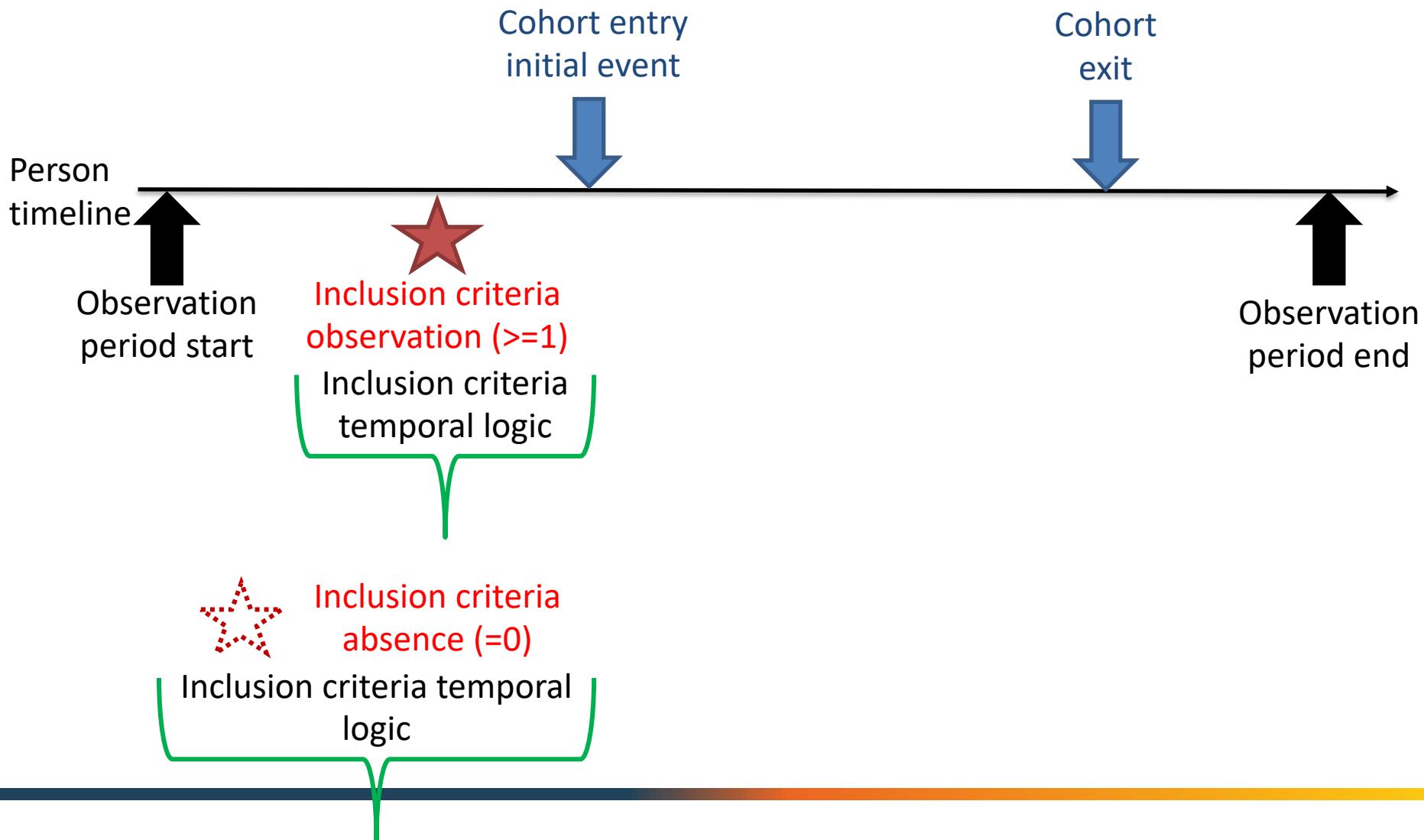
Cohort = a set of persons who satisfy one or more inclusion criteria for a duration of time

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



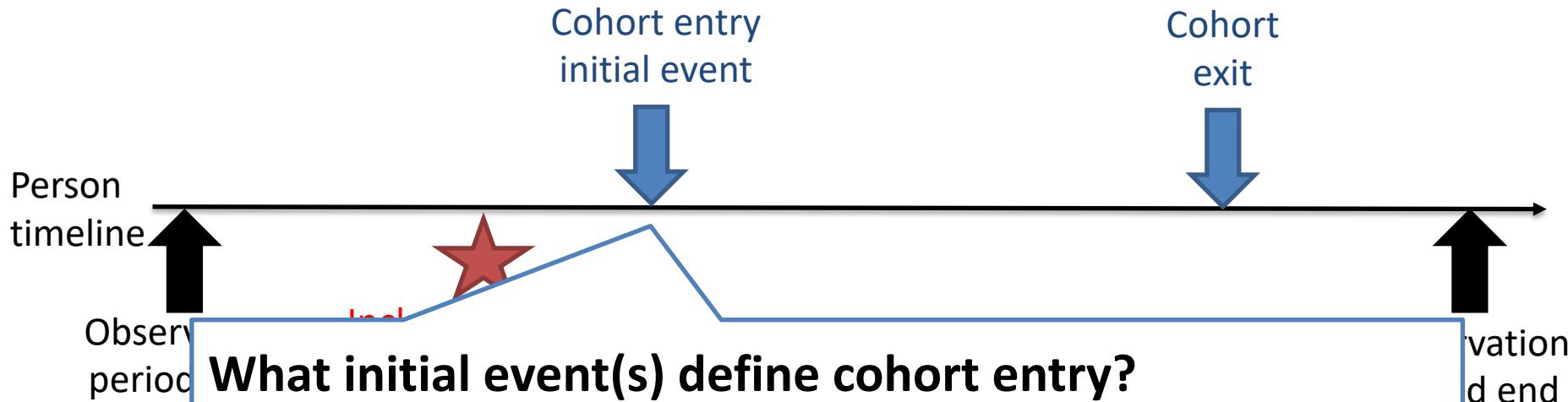


Questions to answer when defining a cohort

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



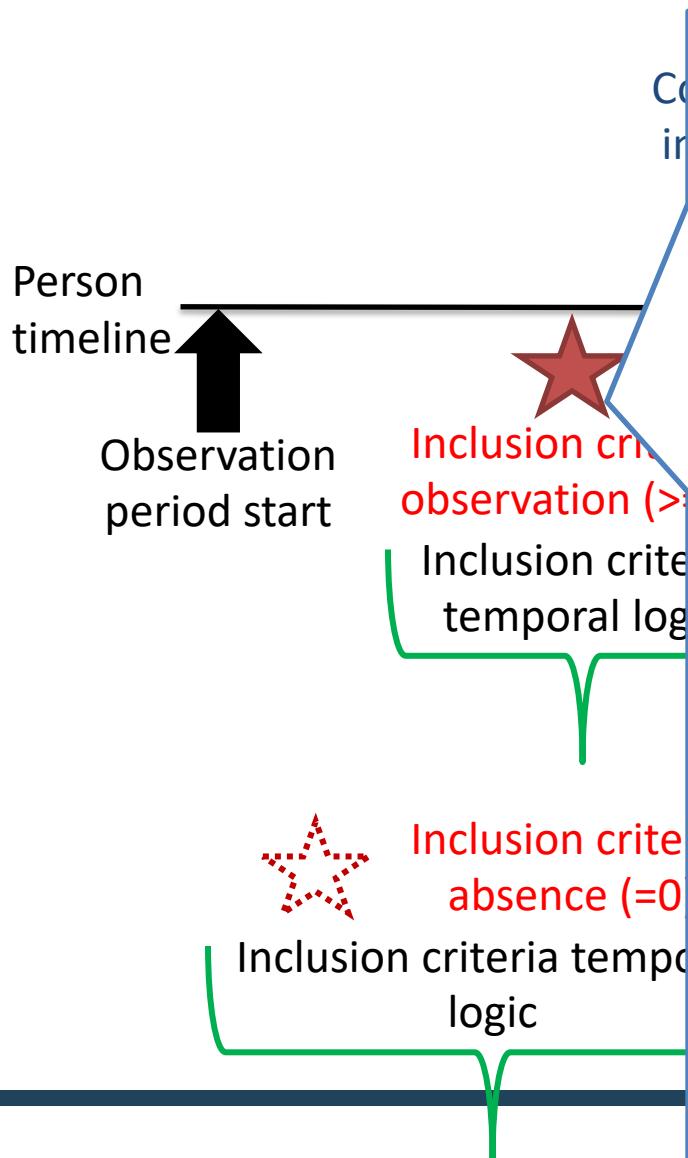
Dissecting the anatomy of a cohort definition



- 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



Dissecting the anatomy of a cohort definition



What inclusion criteria are applied to the initial events?

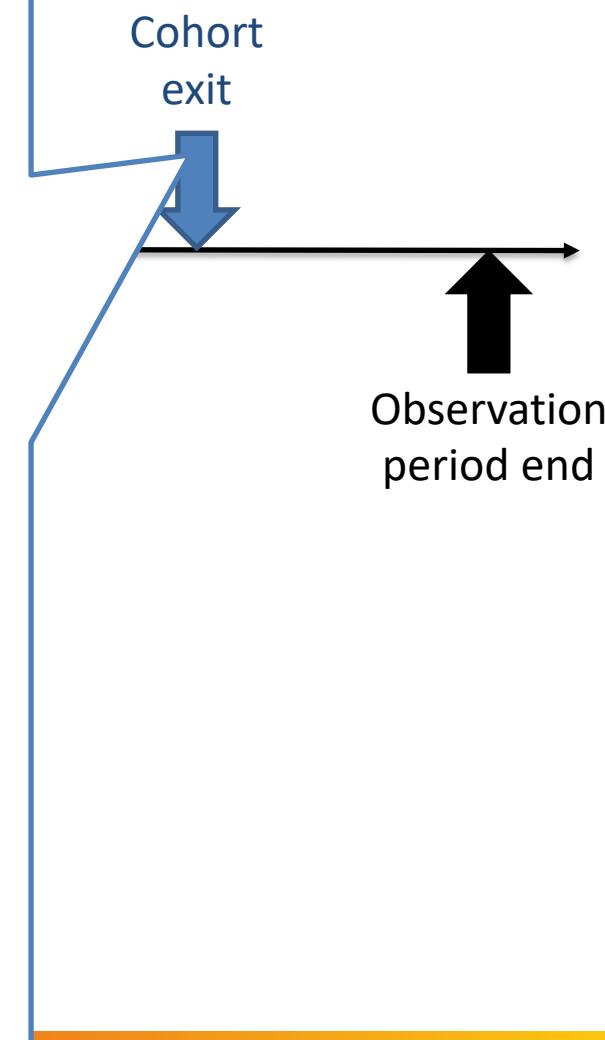
- The qualifying cohort will be defined as all persons who have an initial event and satisfy all qualifying inclusion criteria.
- Each inclusion criteria is defined by domain(s), conceptset(s), domain-specific attributes, and the temporal logic relative to initial events
- Each qualifying inclusion criteria 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)



Dissecting the anatomy of a cohort definition

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. International Classification of Disease, 9th edition, Clinical Modification (ICD 9-CM) codes used to define study outcomes.

| Outcome | ICD-9 Codes | Position | Setting |
|-----------------|---|------------|---------|
| AMI | 410 (all) | 1st or 2nd | IP only |
| Ischemic stroke | 433.x1, 434.x (except subcode: x0), 436 | 1st | IP only |



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?



Graham et al. description of the cohort(s)

A new-user retrospective cohort design was used to compare patients initiating dabigatran or warfarin for the treatment of nonvalvular AF.¹⁰ 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 PheEvaluator

Patrick Ryan

Janssen Research and Development
Columbia University Medical Center



Questions?

Thanks for joining
the journey!

ryan@ohdsi.org