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

Patrick Ryan, PhD

Janssen Research and Development, Columbia University Medical Center

Christian Reich, MD PhD

Iqvia

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. Exploring the vocabulary in ATLAS
3. Designing cohorts
4. Implementing cohort definitions in ATLAS
   - Explore the data to find the lego bricks
   - Put the lego bricks together
5. Evaluating cohort definitions
   - Validation
   - Characterization
6. Build cohorts to support reliable evidence generation

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
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			
ACEI/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. ACEI/ARB indicates angiotensin converting-enzyme inhibitor/angiotensin receptor blocker; NSAIDs, nonsteroidal anti-inflammatory drugs; and SSRI, selective serotonin reuptake inhibitor.

*The $CHADS_2$ 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	
	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

	[redacted]	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 basline 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:

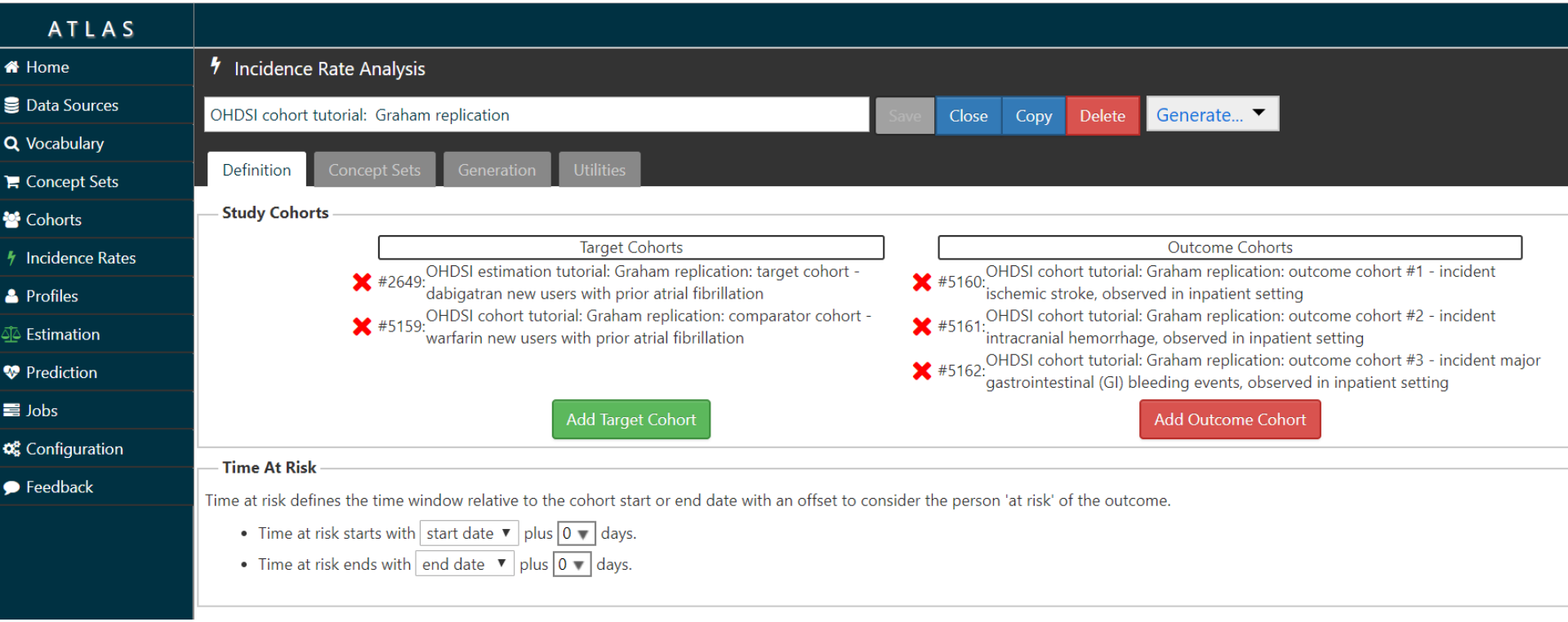
Showing 1 to 15 of 305 entries | Previous 1 2 3 4 5 … 21 Next

Analysis
- Group Era (1025)
- Era (681)

Time Window
- Long Term (708)
- Short Term (537)
- Overlapping (461)

	Concept Name	Time Window	Person Count	% of cohort
Explore	dabigatran etexilate	Long Term	19,975	100.00
Explore	Metoprolol	Long Term	8,820	44.20
Explore	Hydrochlorothiazide	Long Term	5,955	29.90
Explore	Acetaminophen	Long Term	5,739	28.80
Explore	Lisinopril	Long Term	4,935	24.80
Explore	Simvastatin	Long Term	4,851	24.30
Explore	Amlodipine	Long Term	4,808	24.10
Explore	Furosemide	Long Term	4,795	24.10
Explore	Hydrocodone	Long Term	4,590	23.00
Explore	atorvastatin	Long Term	4,422	22.20

Graham replication: Incidence summary design in ATLAS



Graham replication: Incidence summary implementation in ATLAS

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

Export Analysis to CSV

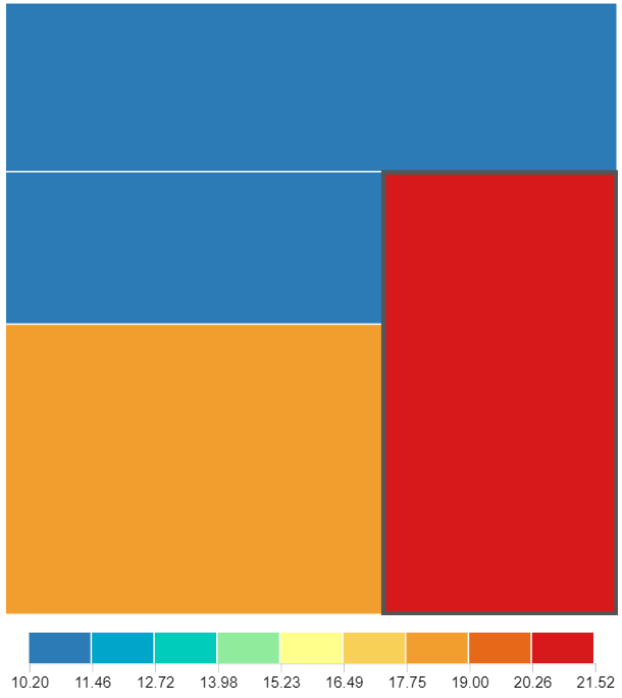
Source	Name		Persons	Cases	Proportion [+|-] per 1k persons	Time At Risk (years)	Rate [+|-] per 1k years	Started	Duration	
TRUVENMDCR_V657	Truven MDCR	Execute	18,376	93	5.06	5,852	15.89	2017-12-02, 22:46	00:00:29	Remove

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

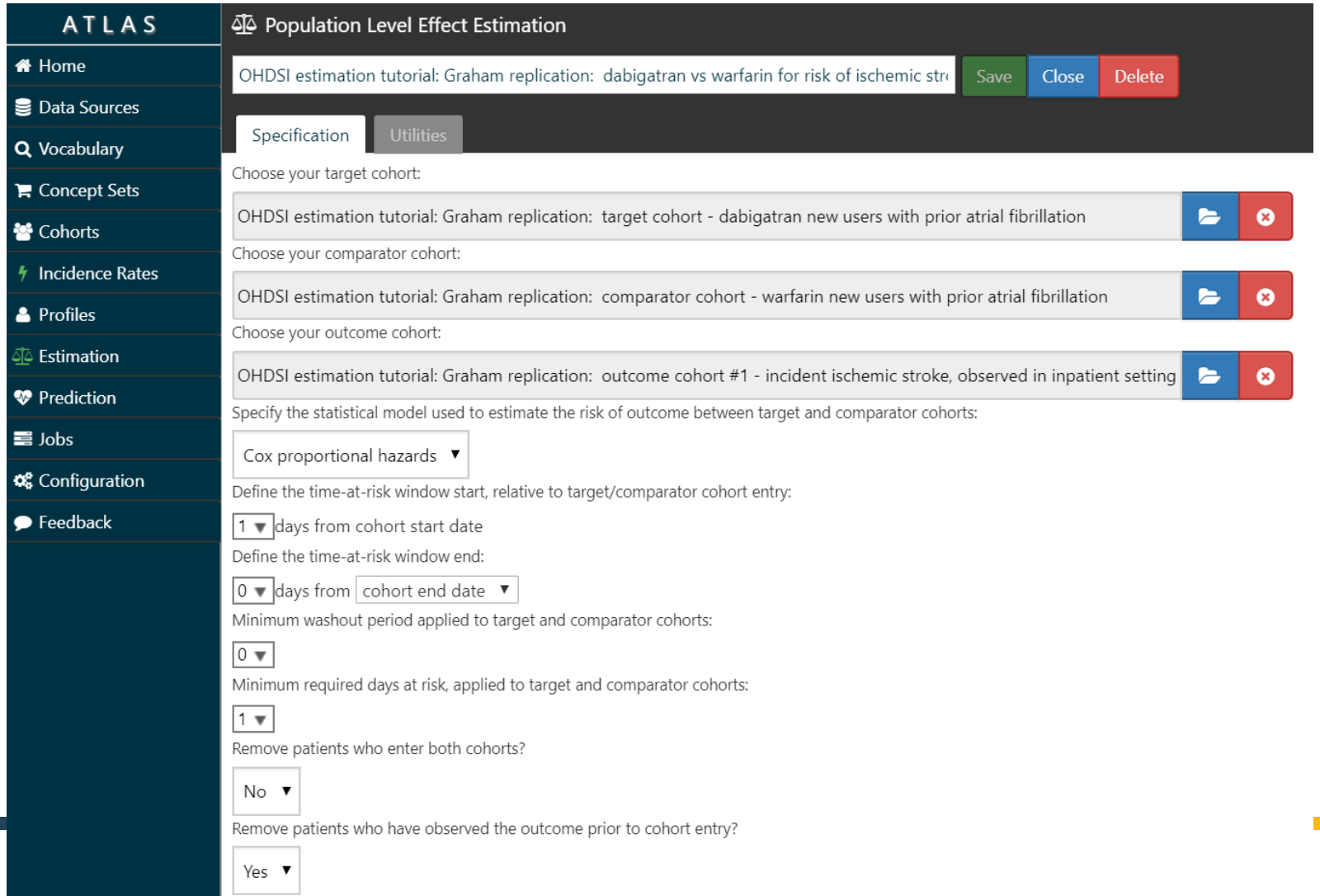
	Persons	Cases	Proportion [+|-] per 1k persons	Time At Risk (years)	Rate [+|-] per 1k years
Summary Statistics:	18,376	93	5.06	5,852	15.89

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

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

Model type: cox
Stratified: FALSE
Use covariates: FALSE
Status: OK

	Estimate	lower .95	upper .95
treatment	0.89626	0.71863	1.11829

Population counts

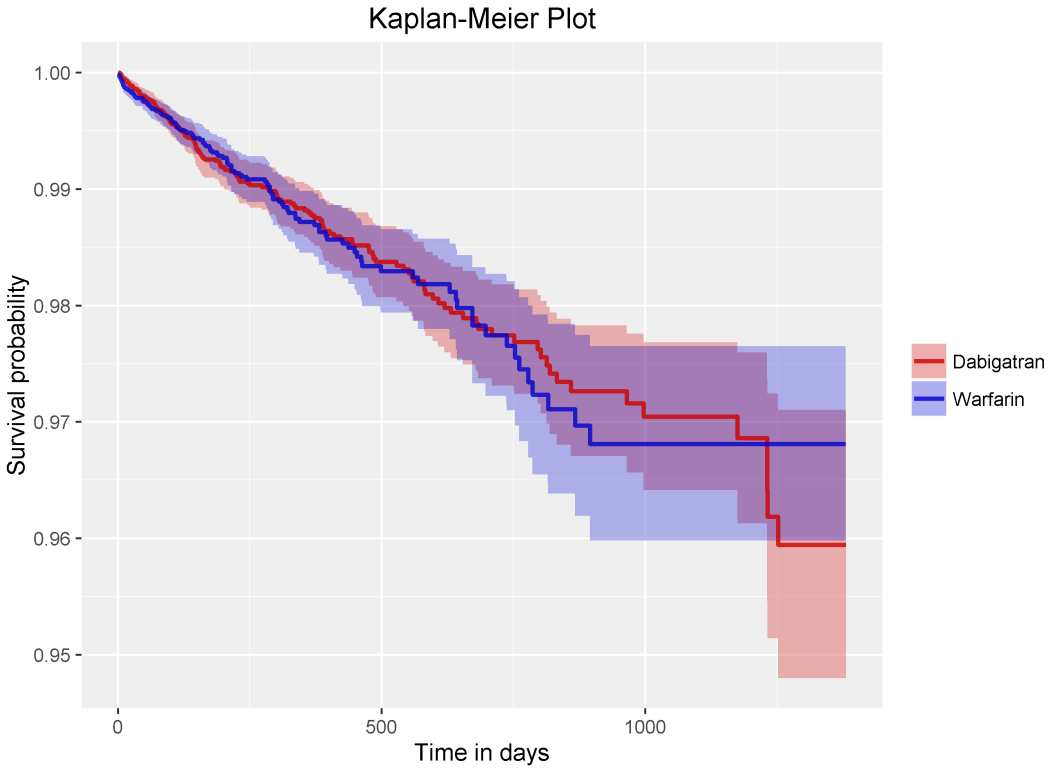
	treatedPersons	comparat
Count	17460	17460

Outcome counts

	treatedPersons	comparat
Count	164	155

Time at risk

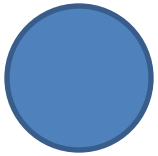
	treatedDays	comparat
Days	4912947	3954046



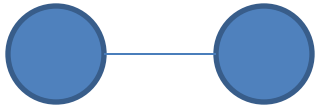
Exploring the vocabulary in ATLAS

http://cohortohdieurope.eu-west-1.elasticbeanstalk.com

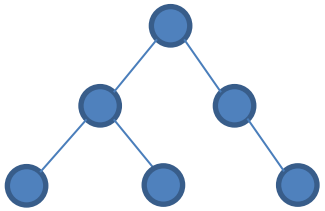
Structure of OMOP Vocabulary



All content: concepts in **`concept`**



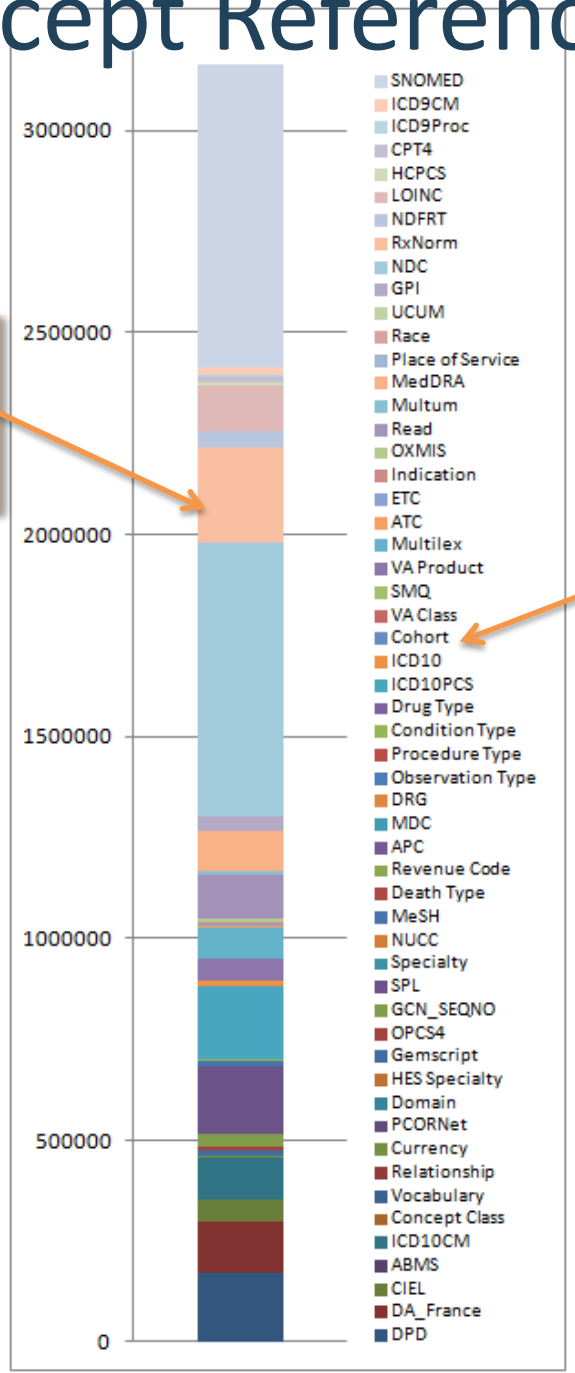
Direct relationships between concepts in **`concept_relationship`**



Multi-step hierarchical relationships pre-processed into **`concept_ancestor`**

Single Concept Reference Table

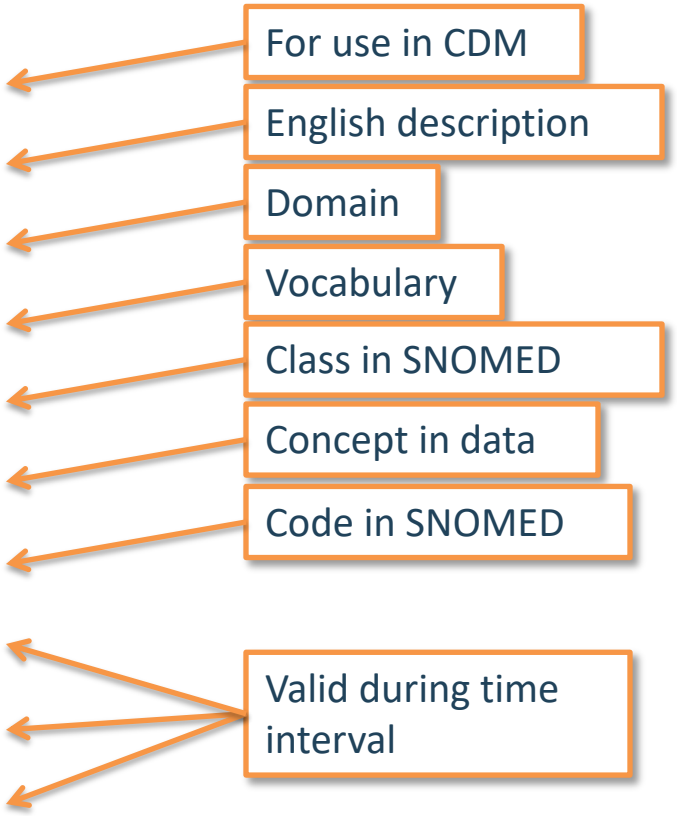
All vocabularies stacked up in one table



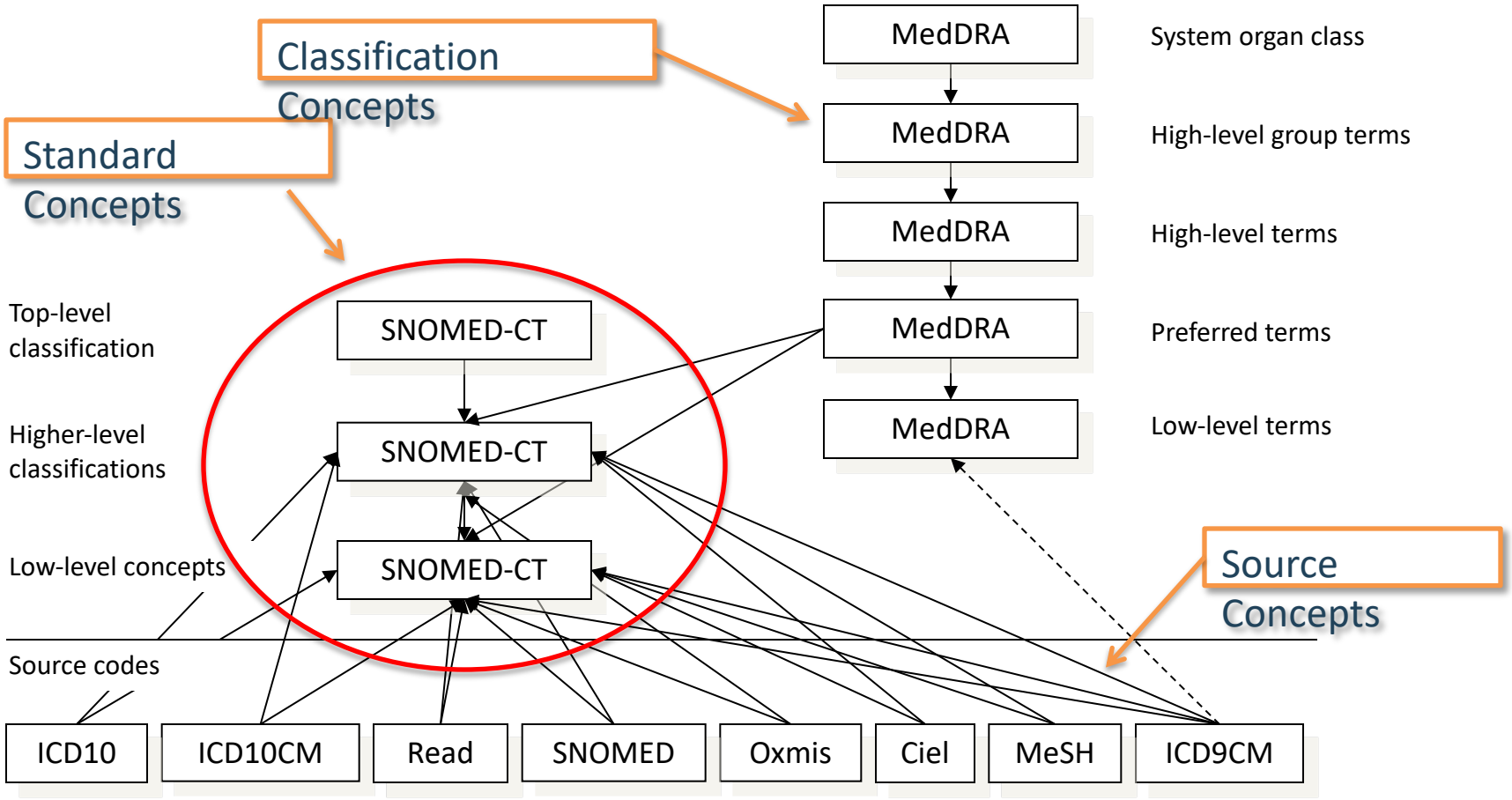
Vocabulary ID

What's in a Concept

CONCEPT_ID	313217
CONCEPT_NAME	Atrial fibrillation
DOMAIN_ID	Condition
VOCABULARY_ID	SNOMED
CONCEPT_CLASS_ID	Clinical Finding
STANDARD_CONCEPT	S
CONCEPT_CODE	49436004
VALID_START_DATE	01-Jan-1970
VALID_END_DATE	31-Dec-2099
INVALID_REASON	



Condition Concepts



Finding the Right Concept #1

1. ..if I know the **ID**

```
SELECT * FROM concept WHERE concept_id = 313217
```

CONCEPT_ID	CONCEPT_ NAME	DOMAIN_ID	VOCABULARY _ID	CONCEPT_ CLASS_ID	STANDARD_ CONCEPT	CONCEPT_ CODE	VALID_START _DATE	VALID_END _DATE	INVALID _REASON
313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	01-Jan-1970	31-Dec-2099	

2. ..if I know the **code**

SNOMED code

```
SELECT * FROM concept WHERE concept_code = '49436004'
```

CONCEPT_ID	CONCEPT_ NAME	DOMAIN_ID	VOCABULARY _ID	CONCEPT_ CLASS_ID	STANDARD_ CONCEPT	CONCEPT_ CODE	VALID_START _DATE	VALID_END _DATE	INVALID _REASON
313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	01-Jan-1970	31-Dec-2099	

Finding the Right Concept #2

3. ..if I know the **name**

```
SELECT * FROM concept WHERE concept_name = 'Atrial fibrillation';
```

CONCEPT_ID	CONCEPT_ NAME	DOMAIN_ID	VOCABULARY_ID	CONCEPT_ CLASS_ID	STANDARD_ CONCEPT	CONCEPT_ CODE
313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004
44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code		427.31
35204953	Atrial fibrillation	Condition	MedDRA	PT	C	10003658
45500085	Atrial fibrillation	Condition	Read	Read		G573000
45883018	Atrial fibrillation	Meas Value	LOINC	Answer	S	LA17084-7

Finding the Right Concept #3

1. if don't know any of this, but I know the code in another vocabulary

ICD-9 is not a Standard Concept

```
SELECT * FROM concept WHERE concept_code = '427.31';
```

CONCEPT _ID	CONCEPT_ NAME	DOMAIN _ID	VOCABULARY _ID	CONCEPT_ CLASS_ID	STANDARD_ CONCEPT	CONCEPT_ CODE
44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code		427.31

```
SELECT * FROM concept_relationship WHERE concept_id_1 = 44821957;
```

Mapping to different vocabularies

Kind of relationship

_ID_1	CONCEPT _ID_2	RELATIONSHIP _ID	VALID_START _DATE	VALID_END _DATE	INVALID _REASON
44821957	21001551	ICD9CM - FDB Ind	01-Oct-13	31-Dec-2099	
44821957	35204953	ICD9CM - MedDRA	01-Jan-70	31-Dec-2099	
44821957	44824248	Is a	01-Oct-14	31-Dec-2099	
44821957	44834731	Is a	01-Oct-14	31-Dec-2099	
44821957	313217	Maps to	01-Jan-70	31-Dec-2099	

Why are we mapping?

European Commission

LANGUAGES

Supporting language learning and linguistic diversity

European Commission › Languages › Policy › Linguistic diversity

Official languages of the EU

What is it?

The European Union has 24 official and working languages. They are:

Bulgarian	French	Maltese
Croatian	German	Polish
Czech	Greek	Portuguese
Danish	Hungarian	Romanian
Dutch	Irish	Slovak
English	Italian	Slovenian
Estonian	Latvian	Spanish
Finnish	Lithuanian	Swedish

What is the Commission doing?

With a permanent staff of 1,750 linguists and 600 support staff, the Commission has one of the largest translation services in the world, bolstered by a further 600 full-time and 3,000 freelance interpreters.

How many different ways do you express one meaning?

Cheers

Mapping = Translating

Step 1. Lookup the Source Concept

```
SELECT * FROM concept WHERE concept_code = '427.31';
```

CONCEPT _ID	CONCEPT_ NAME	DOMAIN _ID	VOCABULARY _ID	CONCEPT_ CLASS_ID	STANDARD_ CONCEPT	CONCEPT_ CODE
44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code		427.31

Step 2. Translate to Standard

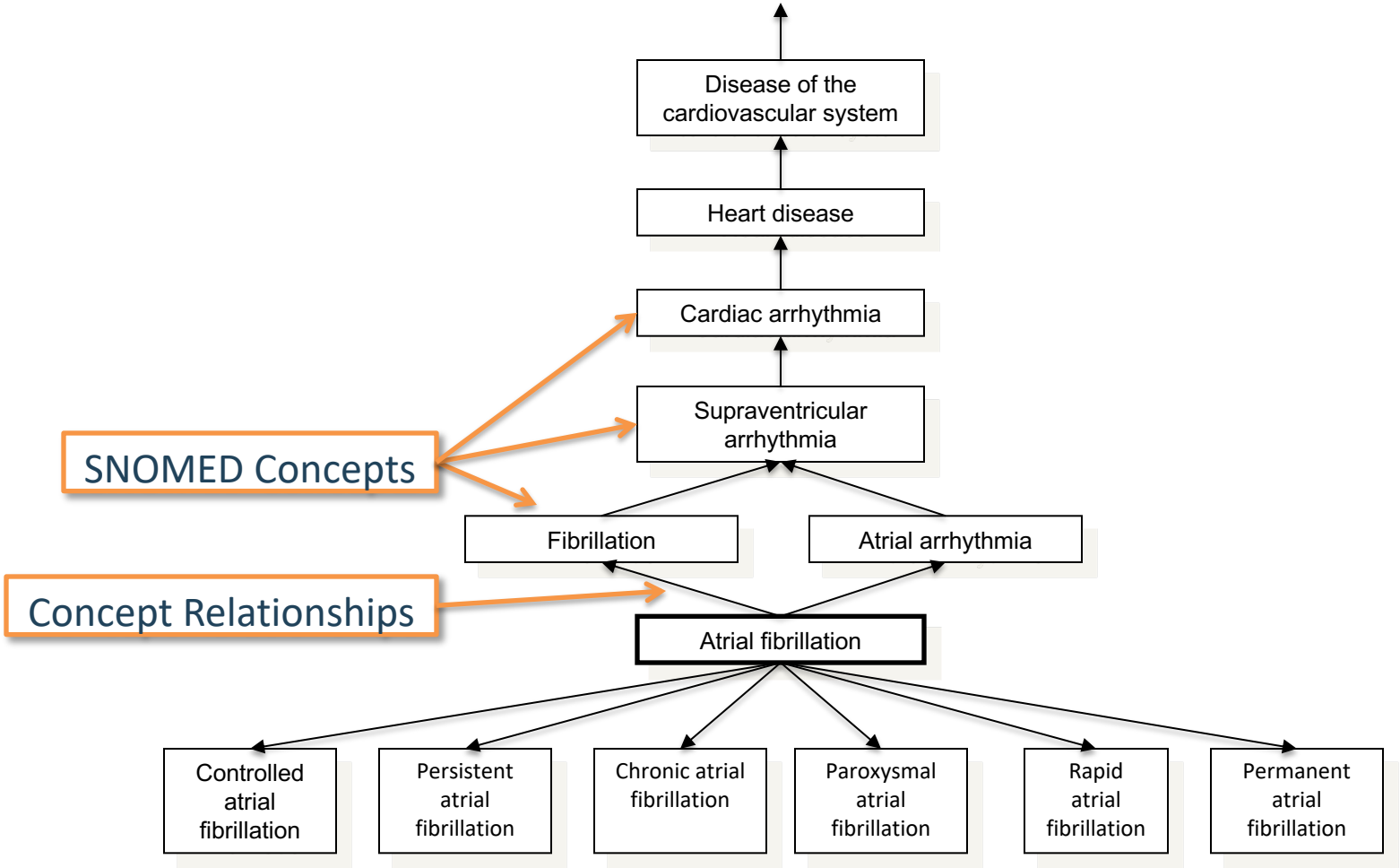
```
SELECT * FROM concept_relationship WHERE concept_id_1 = 44821957
    AND relationship_id = 'Maps to';
```

CONCEPT _ID_1	CONCEPT _ID_2	RELATIONSHIP _ID	VALID_START _DATE	VALID_END _DATE	INVALID _REASON
44821957	313217	Maps to	01-Jan-1970	31-Dec-2099	

Step 3. Check out the translated Concept

```
SELECT * FROM concept WHERE concept_id = 313217;
```

Disease Hierarchy



Exploring Relationships

```
SELECT *
FROM concept_relationship
WHERE concept_id_1 = 313217
```

Related Concepts

Relationship ID

CONCEPT _ID_1	CONCEPT _ID_2	RELATIONSHIP _ID
313217	4232697	Subsumes
313217	4181800	Focus of
313217	35204953	SNOMED - MedDRA eq
313217	4203375	Asso finding of
313217	4141360	Subsumes
313217	4119601	Subsumes
313217	4117112	Subsumes
313217	4232691	Subsumes
313217	4139517	Due to of
313217	4194288	Asso finding of
313217	44782442	Subsumes
313217	44783731	Focus of
313217	21003018	SNOMED - ind/CI
313217	40248987	SNOMED - ind/CI
313217	21001551	SNOMED - ind/CI
313217	21001540	SNOMED - ind/CI
313217	45576876	Mapped from
313217	44807374	Asso finding of
313217	21013834	SNOMED - ind/CI
313217	21001572	SNOMED - ind/CI
313217	21001606	SNOMED - ind/CI
313217	21003176	SNOMED - ind/CI
313217	4226399	Is a
313217	500001801	SNOMED - HOI
313217	500002401	SNOMED - HOI
313217	4119602	Subsumes
313217	40631039	Subsumes
313217	4108832	Subsumes
313217	21013671	SNOMED - ind/CI
313217	21013390	SNOMED - ind/CI
313217	313217	Maps to
313217	44821957	Mapped from
313217	2617597	Mapped from
313217	45500085	Mapped from
313217	313217	Mapped from
313217	45951191	Mapped from
313217	21013856	SNOMED - ind/CI
313217	21001575	SNOMED - ind/CI
313217	21001594	SNOMED - ind/CI

Exploring Relationships

```
SELECT cr.relationship_id, c.*
FROM concept_relationship cr
JOIN concept c ON cr.concept_id_2 = c.concept_id
WHERE cr.concept_id_1 = 313217
```

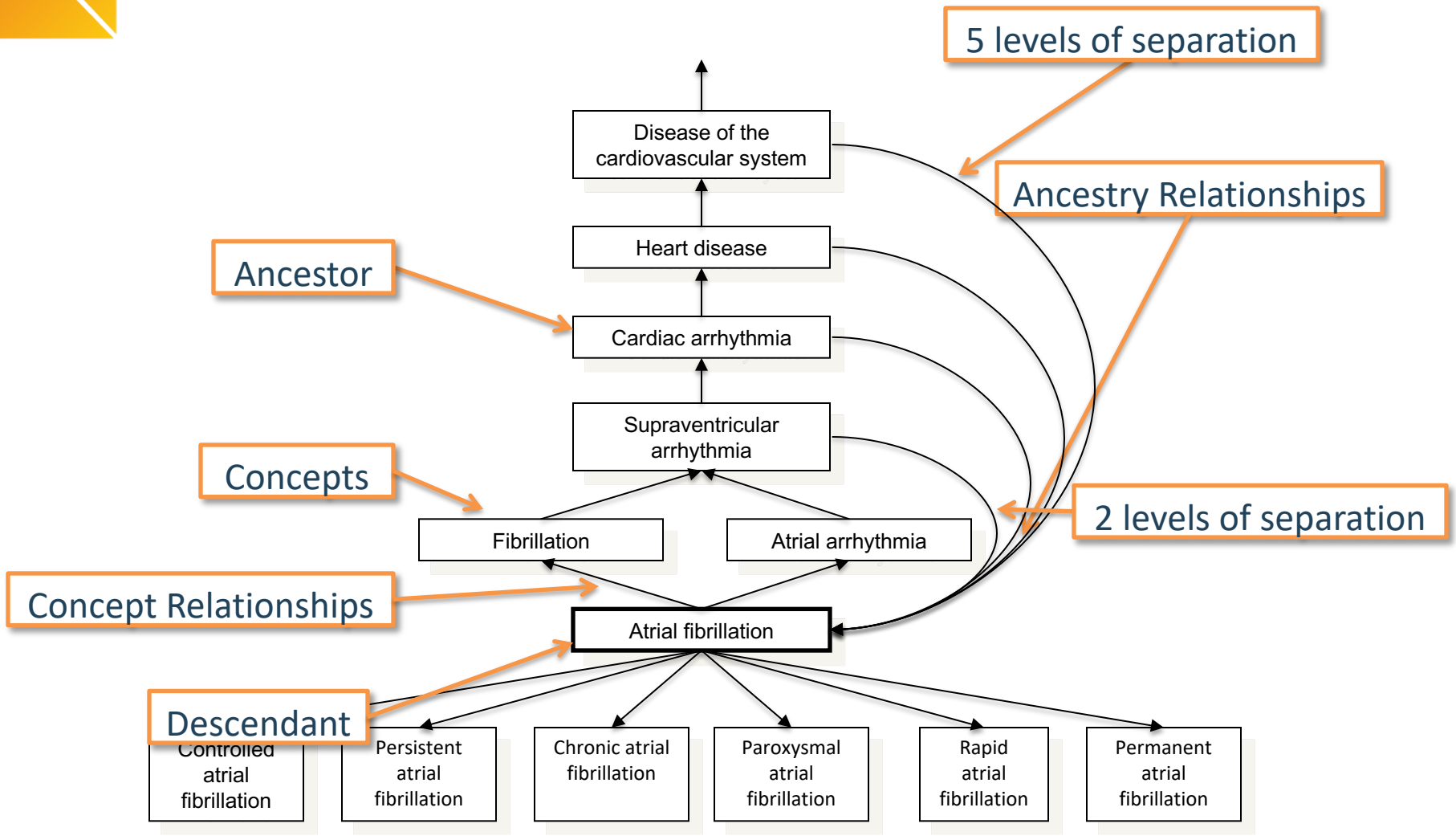
Find out related concept

Ancestor concepts

Descendant concepts

relationship_id	concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date	valid_end_date	invalid_reason
Asso finding of	4194288		Observation	SNOMED	Context-dependent	S	312442005	1/1/1970 0:00	12/31/2099 0:00	NULL
Asso finding of	4203375		Observation	SNOMED	Context-dependent	S	433276002	1/31/2009 0:00	12/31/2099 0:00	NULL
Asso finding of	42689685	Atrial fibrillation not d[illegible]cted	Observation	SNOMED	Context-dependent	S	1.06706E+15	4/1/2017 0:00	12/31/2099 0:00	NULL
Asso finding of	44807374	Atrial fibrillation [illegible]ected	Observation	SNOMED	Context-dependent	S	8.16401E+14	4/1/2014 0:00	12/31/2099 0:00	NULL
Concept poss_eq from	40323929	Fibrillation - atrial	Condition	SNOMED	Clinical Finding	NULL	155364009	1/1/1970 0:00	3/11/2016 0:00	U
Concept poss_eq from	40345197	Fibrillation - atrial	Condition	SNOMED	Clinical Finding	NULL	266306001	1/1/1970 0:00	3/11/2016 0:00	U
Due to of	413952[illegible]	Transient cerebral ischemia due to atrial fibrillation	Condition	SNOMED	Clinical Finding	S	426814001	1/1/1970 0:00	12/31/2099 0:00	NULL
Focus of	4[illegible]709991	Insertion of pacemaker for control of atrial fibrillation	Procedure	SNOMED	Procedure	S	449863006	1/31/2012 0:00	12/31/2099 0:00	NULL
Has finding site	4242112	Atrial structure	Spec Anatomic Site	SNOMED	Body Structure	S	59652004	1/1/1970 0:00	12/31/2099 0:00	NULL
Is a	4226399	Fibrillation	Condition	SNOMED	Clinical Finding	S	40593004	1/1/1970 0:00	12/31/2099 0:00	NULL
Is a	4068155	Atrial arrhythmia	Condition	SNOMED	Clinical Finding	S	17366009	1/1/1970 0:00	12/31/2099 0:00	NULL
Mapped from	40323929	Fibrillation - atrial	Condition	SNOMED	Clinical Finding	NULL	155364009	1/1/1970 0:00	3/11/2016 0:00	U
Mapped from	2617597	Patient with heart failure and atrial fibrillation documented to be on warfarin therapy	Observation	HCPCS	HCPCS	NULL	G8183	1/1/1970 0:00	11/11/2014 0:00	D
Mapped from	45576876	Unspecified atrial fibrillation	Condition	ICD10CM	5-char billing code	NULL	I48.91	12/30/2006 0:00	12/31/2099 0:00	NULL
Mapped from	45500085	Atrial fibrillation	Condition	Read	Read	NULL	G573000	1/1/1970 0:00	12/31/2099 0:00	NULL
Mapped from	45611600	Atrial Fibrillation	Condition	MeSH	Main Heading	NULL	D001281	1/1/1970 0:00	12/31/2099 0:00	NULL
Mapped from	40345197	Fibrillation - atrial	Condition	SNOMED	Clinical Finding	NULL	266306001	1/1/1970 0:00	3/11/2016 0:00	U
Mapped from	45951191	Atrial Fibrillation	Condition	CIEL	Diagnosis	NULL	148203	11/3/2007 0:00	12/31/2099 0:00	NULL
Mapped from	313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	1/1/1970 0:00	12/31/2099 0:00	NULL
Mapped from	44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code	NULL	427.31	1/1/1970 0:00	12/31/2099 0:00	NULL
Maps to	313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - HOI	500002401	OMOP Atrial Fibrillation 1	Condition	Cohort	Cohort	C	500002401	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - HOI	500001801	OMOP Qt Prolongation/Torsade De Pointes 1	Condition	Cohort	Cohort	C	500001801	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - ind/CI	21005673	Prevention of Thromboembolism in Chronic Atrial Fibrillation	Drug	Indication	Indication	C	5673	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - ind/CI	21003176	Tachyarrhythmia	Drug	Indication	Indication	C	3176	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - ind/CI	21001542	Supraventricular Tachycardia	Drug	Indication	Indication	C	1542	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - ind/CI	21001594	Disease of Cardiovascular System	Drug	Indication	Indication	C	1594	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - MedDRA eq	35204953	Atrial fibrillation	Condition	MedDRA	PT	C	10003658	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	4117112	Controlled atrial fibrillation	Condition	SNOMED	Clinical Finding	S	300996004	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	4119601	Lone atrial fibrillation	Condition	SNOMED	Clinical Finding	S	233910005	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	4232697	Persistent atrial fibrillation	Condition	SNOMED	Clinical Finding	S	440059007	1/31/2009 0:00	12/31/2099 0:00	NULL
Subsumes	4141360	Chronic atrial fibrillation	Condition	SNOMED	Clinical Finding	S	426749004	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	44782442	Atrial fibrillation with rapid ventricular response	Condition	SNOMED	Clinical Finding	S	1.20041E+14	1/31/2014 0:00	12/31/2099 0:00	NULL
Subsumes	4199501	Rapid atrial fibrillation	Condition	SNOMED	Clinical Finding	S	314208002	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	4119602	Non-rheumatic atrial fibrillation	Condition	SNOMED	Clinical Finding	S	233911009	1/1/1970 0:00	12/31/2099 0:00	NULL

Ancestry Relationships: Higher-Level Relationships



Exploring Ancestors of a Concept

```
SELECT max_levels_of_separation, concept.*
FROM concept_ancestor
JOIN concept ON ancestor_concept_id = concept_id
WHERE descendant_concept_id = 313217 /* Atrial fibrillation */
ORDER BY max_levels_of_separation
```

max_levels_of_separation	concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept
0	313217	Atrial fibrillation	Co			
0	35204953	Atrial fibrillation	Co			
1	4226399	Fibrillation	Condition	SNOMED	Clinical Finding	S
1	4068155	Atrial arrhythmia	Condition	SNOMED	Clinical Finding	S
1	35204969	Cardiac fibrillation	Condition	MedDRA	PT	C
2	4248028	Supraventricular arrhythmia	Condition	SNOMED	Clinical Finding	S
2	35204952	Arrhythmia supraventricular	Condition	MedDRA	PT	C
2	35202454	Rate and rhythm disorders NEC	Condition	MedDRA	HLT	C
3	44784217	Cardiac arrhythmia	Condition	SNOMED	Clinical Finding	S
3	35202455	Supraventricular arrhythmias	Condition	MedDRA	HLT	C
4	321588	Heart disease	Condition	SNOMED	Clinical Finding	S
4	35204989	Cardiac disorder	Condition	MedDRA	PT	C
4	35202050	Cardiac arrhythmias	Condition	MedDRA	HLGT	C
5	4103183	Cardiac finding	Condition	SNOMED	Clinical Finding	S
5	440142	Disorder of mediastinum	Condition	SNOMED	Clinical Finding	S
5	134057	Disorder of cardiovascular system	Condition	SNOMED	Clinical Finding	S
5	35204998	Cardiovascular disorder	Condition	MedDRA	PT	C
5	37219970	Mediastinal disorder	Condition	MedDRA	PT	C
5	37622411	Phlebosclerosis	Condition	MedDRA	PT	C
5	35202457	Cardiac disorders NEC	Condition	MedDRA	HLT	C
6	4115390	Mediastinal finding	Condition	SNOMED	Clinical Finding	S
6	4023995	Cardiovascular finding	Condition	SNOMED	Clinical Finding	S

Hold the descendant

Exploring Descendants of a Concept

```
SELECT max_levels_of_separation, concept.*
FROM concept_ancestor
JOIN concept ON descendant_concept_id = concept_id
WHERE ancestor_concept_id = 44784217 /* cardiac arrythmia */
ORDER BY max_levels_of_separation
```

Hold the ancestor

MAX_LEVELS_OF_SEPARATION	CONCEPT_ID	CONCEPT_NAME	DOMAIN_ID	VOCABULARY_ID	CONCEPT_CLASS_ID	STANDARD_CONCEPT
0	44784217	Cardiac arrhythmia	Condition	SNOMED	Clinical Finding	S
1	313224	Anomalous atrioventricular excitation	Condition	SNOMED	Clinical Finding	S
1	315643	Tachyarrhythmia	Condition	SNOMED	Clinical Finding	S
1	316429	Premature beats	Condition	SNOMED	Clinical Finding	S
1	316999	Conduction disorder of the heart	Condition	SNOMED	Clinical Finding	S
1	321042	Cardiac arrest	Condition	SNOMED	Clinical Finding	S
1	4030583	Pacemaker twiddler's syndrome	Condition	SNOMED	Clinical Finding	S
1	4057008	Accelerated atrioventricular conduction	Condition	SNOMED	Clinical Finding	S
1	4086313	Withdrawal arrhythmia	Condition	SNOMED	Clinical Finding	S
1	4088507	Ventricular escape complex	Condition	SNOMED	Clinical Finding	S
1	4088986	Atrial escape complex	Condition	SNOMED	Clinical Finding	S
1	4091901	Aberrant premature complexes	Condition	SNOMED	Clinical Finding	S
1	4092011	Aberrantly conducted complex	Condition	SNOMED	Clinical Finding	S
1	4124704	Postoperative sinoatrial disease	Condition	SNOMED	Clinical Finding	S
1	4143042	Ectopic beats	Condition	SNOMED	Clinical Finding	S
1	4164083	Ectopic rhythm	Condition	SNOMED	Clinical Finding	S
1	4172863	Fetal dysrhythmia	Condition	SNOMED	Clinical Finding	S
1	4173170	Neonatal dysrhythmia	Condition	SNOMED	Clinical Finding	S
1	4175473	Atrioventricular dissociation	Condition	SNOMED	Clinical Finding	S
1	4185572	Ventricular arrhythmia	Condition	SNOMED	Clinical Finding	S
1	4217221	Nodal rhythm disorder	Condition	SNOMED	Clinical Finding	S
1	4226399	Fibrillation	Condition	SNOMED	Clinical Finding	S
1	4228448	Bradyarrhythmia	Condition	SNOMED	Clinical Finding	S

Let Us find Upper Gastrointestinal Bleeding

1. Find some initiation concept

```
SELECT * FROM concept WHERE concept_name = 'Upper gastrointestinal bleeding'
```

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code
42891225	Upper gastrointest...	Condition	MedDRA	LLT	C	10071910

2. Find standard concepts

```
SELECT * FROM concept WHERE lower(concept_name) LIKE '%upper gastrointestinal%'
    AND domain_id = 'Condition' AND standard_concept = 'S'
```

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code
4308202	Acute upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	38938002
4291649	Upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	37372002
4115581	Finding of upper gastrointestinal gas	Condition	SNOMED	Clinical Finding	S	300370006
4103011	Chronic upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	25349007
4012503	Excessive upper gastrointestinal gas	Condition	SNOMED	Clinical Finding	S	162076009
4000609	Disorder of upper gastrointestinal tract	Condition	SNOMED	Clinical Finding	S	119291004
4332645	Upper gastrointestinal hemorrhage associated with hypercoagulability state	Condition	SNOMED	Clinical Finding	S	430349003

Going up the hierarchy: Finding the right concept

```
SELECT max_levels_of_separation, concept.*
FROM concept_ancestor
JOIN concept ON ancestor_concept_id = concept_id
WHERE descendant_concept_id = 4332645 /* Upper gastrointestinal hemorrhage associated...*/
ORDER BY max_levels_of_separation
```

Hold the descendant

max_levels_of_separation	concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code
0	4332645	Upper gastrointestinal hemorrhage associated with hypercoag	Condition	SNOMED	Clinical Finding	S	430349003
1	35708054	Gastritis haemorrhagic	Condition	MedDRA	PT	C	10017866
1	4291649	Upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	37372002
1	35707871	Upper gastrointestinal haemorrhage	Condition	MedDRA	PT	C	10046274
2	35707864	Gastrointestinal haemorrhage	Condition	MedDRA	PT	C	10017955
2	4000609	Disorder of upper gastrointestinal tract	Condition	SNOMED	Clinical Finding	S	119291004
2	35707858	Intestinal haemorrhage	Condition	MedDRA	PT	C	10059175
2	35702752	Gastritis (excl infective)	Condition	MedDRA	HLT	C	10017854
2	192671	Gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	74474003
3	37604042	Gastrointestinal haemorrhages	Condition	MedDRA	HLT	C	10052742
3	37622518	Haemorrhage	Condition	MedDRA	PT	C	10055798
3	437312	Bleeding	Condition	SNOMED	Clinical Finding	S	131148009
3	4198525	Disorder of upper digestive tract	Condition	SNOMED	Clinical Finding	S	50410009
3	37622515	Extravasation blood	Condition	MedDRA	PT	C	10015867
3	4000610	Disorder of gastrointestinal tract	Condition	SNOMED	Clinical Finding	S	119292006
3	35702116	Gastrointestinal inflammatory conditions	Condition	MedDRA	HLGT	C	10017969
3	35702743	Intestinal haemorrhages	Condition	MedDRA	HLT	C	10022653
3	35702744	Non-site specific gastrointestinal haemorrhages	Condition	MedDRA	HLT	C	10017958
4	35702114	Gastrointestinal haemorrhages NEC	Condition	MedDRA	HLGT	C	10017959
4	4304916	Gastrointestinal tract finding	Condition	SNOMED	Clinical Finding	S	386618008
4	35702767	Nausea and vomiting symptoms	Condition	MedDRA	HLT	C	10028817

Going down the hierarchy : Checking the right content

```
SELECT max_levels_of_separation, concept.*
FROM concept_ancestor
JOIN concept ON descendant_concept_id = concept_id
WHERE ancestor_concept_id = 4291649 /* Upper gastrointestinal hemorrhage */
ORDER BY max_levels_of_separation
```

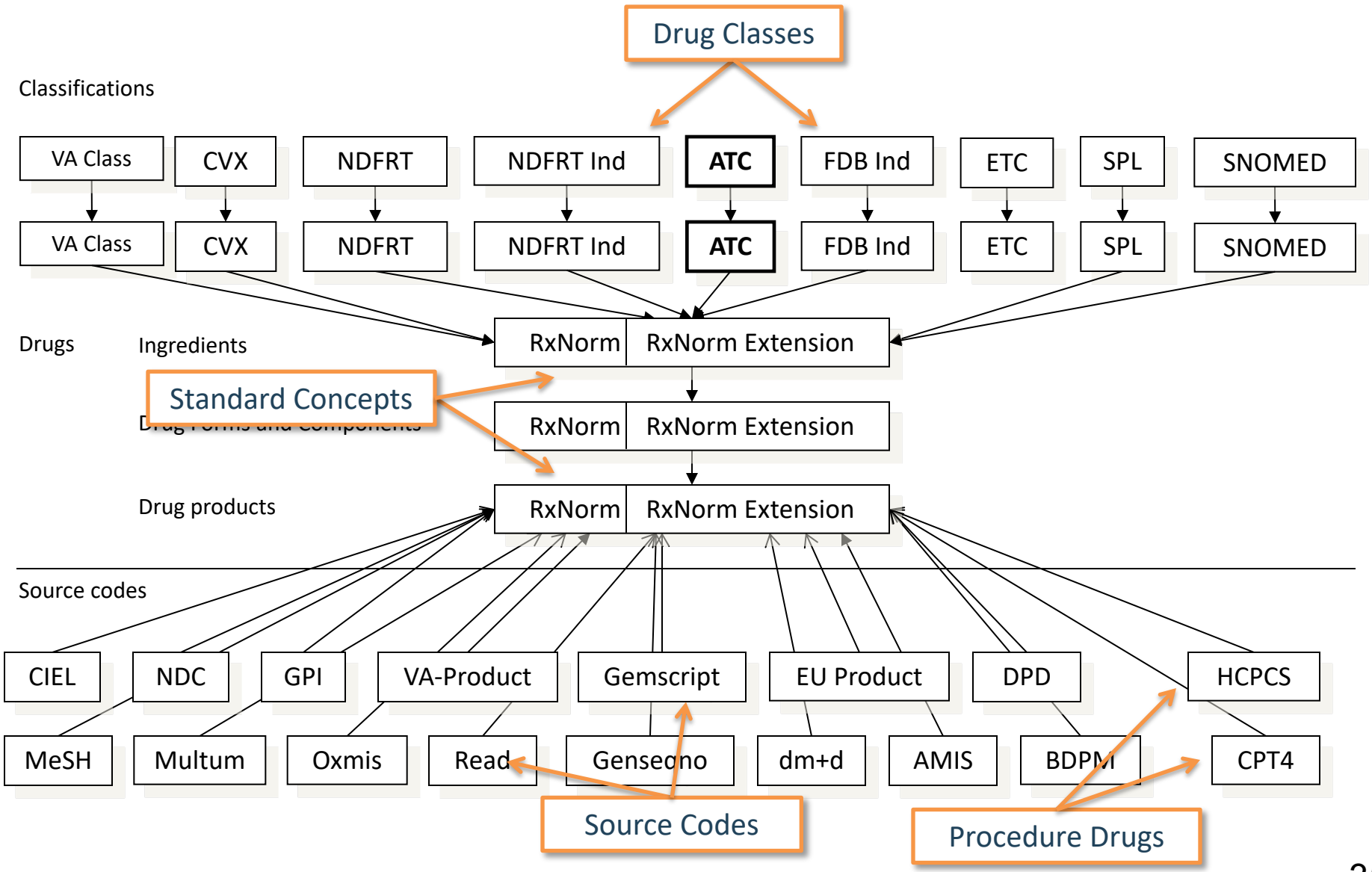
max_levels_of_separation	concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code
0	4291649	Upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	37372002
1	4318535	Duodenal hemorrhage					
1	23245	Esophageal bleeding					
1	4308202	Acute upper gastrointestinal hemorrhage					
1	4271696	Peptic ulcer with hemorrhage	Condition	SNOMED	Clinical Finding	S	64121000
1	4103011	Chronic upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	25349007
1	26727	Hematemesis	Condition	SNOMED	Clinical Finding	S	8765009
1	4332645	Upper gastrointestinal hemorrhage associated with hypercoag	Condition	SNOMED	Clinical Finding	S	430349003
1	193250	Gastric hemorrhage	Condition	SNOMED	Clinical Finding	S	61401005
2	4131525	Hemorrhagic gastropathy	Condition	SNOMED	Clinical Finding	S	413218001
2	4204041	Hematemesis - cause unknown	Condition	SNOMED	Clinical Finding	S	308904008
2	4134808	Hemorrhagic duodenopathy	Condition	SNOMED	Clinical Finding	S	413212000
2	4260059	Hemorrhagic gastroenteritis	Condition	SNOMED	Clinical Finding	S	409506009
2	4099014	Duodenal ulcer with hemorrhage	Condition	SNOMED	Clinical Finding	S	27281001
2	46270145	Gastric hemorrhage due to atrophic gastritis	Condition	SNOMED	Clinical Finding	S	1.5072E+14
2	4096032	Duodenal hematoma	Condition	SNOMED	Clinical Finding	S	262843005
2	4174044	Chronic peptic ulcer with hemorrhage	Condition	SNOMED	Clinical Finding	S	49232000
2	4095555	Esophageal hematoma	Condition	SNOMED	Clinical Finding	S	262790002
2	46269904	Hemorrhage of duodenum co-occurrent and due to diverticulo	Condition	SNOMED	Clinical Finding	S	1.0866E+15
2	45768629	Gastric hemorrhage due to erosive gastritis	Condition	SNOMED	Clinical Finding	S	7.071E+12

Concept 4291649 and all its descendants comprise Upper GI Bleeding

Does it Work that Way with Drugs?

- Codes
  - NDC, GPI, Multilex, HCPCS, etc.
- Concepts
  - Drug products (Generic and Brand)
  - Drug ingredients
  - Drug Classes
- Relationships
- Ancestry

Drug Hierarchy



Let us find Warfarin

1. Find active compound Warfarin by keyword

```
SELECT * FROM concept WHERE lower(concept_name) = 'warfarin'
```

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code
21253542	Warfarin	Drug	dm+d	VTM	NULL	48603004
40772658	Warfarin	Measurement	LOINC	LOINC Hierarchy	C	LP16309-4
1847834	WARFARIN	Drug	DA_France	Ingredient	NULL	OMOP13547
4293218	WARFARIN	Drug	NDFRT	Pharma Preparation	NULL	N0000148057
35715898	WARFARIN	Drug	LPD_Australia	Ingredient	NULL	OMOP582219
43081820	warfarin	Drug	Multilex	Ingredient	NULL	2849
4187015	Warfarin	Drug	SNOMED	Substance	NULL	372756006
43343324	Warfarin	Drug	AMT	AU Substance	NULL	2714011000036109
4174989	Warfarin	Drug	SNOMED	Pharma/Biol Product	NULL	48603004
4325514	Warfarin	Drug	NDFRT	Chemical Structure	C	N0000006403
1310149	Warfarin	Drug	RxNorm	Ingredient	S	11289
21600965	warfarin	Drug	ATC	ATC 5th	C	B01AA03
45618204	Warfarin	Drug	MeSH	Main Heading	NULL	D014859

Let us find Clopidogrel

1. Find drug product containing Clopidogrel by NDC code:

   Bristol Meyer Squibb's Plavix 75mg capsules: NDC 67544050474

```
SELECT * FROM concept WHERE concept_code = '67544050474'
```

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date	valid_end_date	invalid_reason
45867731	clopidogrel 75 MG Oral Tablet [Plavix]	Drug	NDC	11-digit NDC	NULL	67544050474	2014-07-01	2099-12-31	NULL

```
SELECT * FROM concept_relationship WHERE concept_id_1 = 45867731
 AND relationship_id = 'Maps to'
```

concept_id_1	concept_id_2	relationship_id	valid_start_date	valid_end_date	invalid_reason
45867731	1322185	Maps to	2015-01-29	2099-12-31	NULL

```
SELECT * FROM concept WHERE concept_id = 1322185
```

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date	valid_end_date	invalid_reason
1322185	clopidogrel 75 MG Oral Tablet [Plavix]	Drug	RxNorm	Branded Drug	S	213169	1970-01-01	2099-12-31	NULL

Let us find Clopidogrel ingredient

2. Find ingredient Clopidogrel as Ancestor of drug product

```
SELECT max_levels_of_separation, concept.*
FROM concept_ancestor
JOIN concept ON ancestor_concept_id = concept_id
WHERE descendant_concept_id = 1322185 /* clopidogrel 75 MG Oral Tablet [Plavix]*/
ORDER BY max_levels_of_separation
```

max_levels_of_separation	concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code
0	1322185	clopidogrel 75 MG Oral Tablet [Plavix]	Drug	RxNorm	Branded Drug	S	213169
0	19075601	clopidogrel 75 MG Oral Tablet	Drug	RxNorm	Clinical Drug	S	309362
1	40095879	clopidogrel Oral Tablet [Plavix]	Drug	RxNorm	Branded Drug Form	S	368301
1	19120256	clopidogrel 75 MG [Plavix]	Drug	RxNorm	Branded Drug Comp	S	573094
1	1322187	clopidogrel 75 MG	Drug	RxNo		S	329449
1	40095878	clopidogrel Oral Tablet	Drug	RxNo		S	374583
2	36222254	clopidogrel Oral Product	Drug	RxNorm		C	1163766
2	36229332	Plavix Pill	Drug	RxNorm	Branded Dose Group	C	1181791
2	36229331	Plavix Oral Product	Drug	RxNorm	Branded Dose Group	C	1181790
2	36222255	clopidogrel Pill	Drug	RxNorm	Clinical Dose Group	C	1163767
2	1322184	clopidogrel	Drug	RxNorm	Ingredient	S	32968
3	46319141	CLOPIDOGREL - clopidogrel tablet, film coated	Drug	SPL	Prescription Drug	C	52adfb2c-2062-495c-9954-39eeecae2b41
3	4279519	PLATELET AGGREGATION INHIBITORS	Drug	VA Class	VA Class	C	BL117
3	45796809	clopidogrel 75mg/1 ORAL TABLET, FILM COATED [clopidogrel b	Drug	SPL	Prescription Drug	C	b4e53c96-e280-47c6-baa0-ec676e041d8d
3	45798740	clopidogrel bisulfate 75mg/1 ORAL TABLET, FILM COATED	Drug	SPL	Prescription Drug	C	c7fa330d-d8f1-487e-a730-bafae123e9a8
3	21600985	Platelet aggregation inhibitors excl. heparin	Drug	ATC		C	B01AC

Clopidogrel

Drug classes

Check out Ingredients

3. Check Descendants (other drug products containing Warfarin and Dabigatran)

```
SELECT max_levels_of_separation, concept.*
FROM concept_ancestor
JOIN concept ON descendant_concept_id = concept_id
WHERE ancestor_concept_id = 1310149 /* Warfarin or 1322185 Clopidogrel*/
ORDER BY max_levels_of_separation
```

concept_id	concept_name	vocabulary_id	concept_class_id
1310149	Warfarin	RxNorm	Ingredient
36221229	Jantoven Pill	RxNorm	Branded Dose Group
40163559	Warfarin Sodium 6 MG	RxNorm	Clinical Drug Comp
40163544	Warfarin Sodium 3 MG [Jantoven]	RxNorm	Branded Drug Comp
21134746	Warfarin 0.2 MG/ML	RxNorm Extension	Clinical Drug Comp
21105414	Warfarin 5 MG/ML	RxNorm Extension	Clinical Drug Comp
36221228	Jantoven Oral Product	RxNorm	Branded Dose Group
40163565	Warfarin Sodium 7.5 MG	RxNorm	Clinical Drug Comp
21115236	Warfarin 0.3 MG/ML	RxNorm Extension	Clinical Drug Comp
40163509	Warfarin Sodium 1 MG	RxNorm	Clinical Drug Comp
21156284	1 ML Warfarin 0.02 MG/ML Oral Solution	RxNorm Extension	Quant Clinical Drug
21095537	Warfarin 0.3 MG/ML Oral Solution	RxNorm Extension	Clinical Drug
21105427	Warfarin 0.4 MG/ML Oral Solution	RxNorm Extension	Clinical Drug
21046557	Warfarin 1 MG/ML Oral Solution	RxNorm Extension	Clinical Drug
40093133	Warfarin Oral Tablet [Coumadin]	RxNorm	Branded Drug Form
40093134	Warfarin Oral Tablet [Jantoven]	RxNorm	Branded Drug Form
21077698	1 ML Warfarin 1 MG/ML Oral Solution	RxNorm Extension	Quant Clinical Drug
40163534	Warfarin Sodium 2.5 MG Oral Tablet	RxNorm	Clinical Drug
40163530	Warfarin Sodium 2 MG/ML Injectable Solution	RxNorm	Clinical Drug
21066136	Warfarin 5 MG Oral Tablet [Marevan]	RxNorm Extension	Branded Drug
40163542	Warfarin Sodium 3 MG Oral Tablet [Jantoven]	RxNorm	Branded Drug
21116822	1 ML Warfarin 0.6 MG/ML Oral Suspension	RxNorm Extension	Quant Clinical Drug
21175784	1 ML Warfarin 0.1 MG/ML Oral Solution	RxNorm Extension	Quant Clinical Drug
21175783	1 ML Warfarin 0.832 MG/ML Oral Solution	RxNorm Extension	Quant Clinical Drug

concept_id	concept_name	vocabulary_id	concept_class_id
1322184	clopidogrel	RxNorm	Ingredient
21043471	clopidogrel Oral Suspension	RxNorm Extension	Clinical Drug Form
36229332	Plavix Pill	RxNorm	Branded Dose Group
21043470	clopidogrel Oral Solution	RxNorm Extension	Clinical Drug Form
21023802	clopidogrel Injectable Solution	RxNorm Extension	Clinical Drug Form
21023806	clopidogrel 5 MG	RxNorm Extension	Clinical Drug Comp
1322187	clopidogrel 75 MG	RxNorm	Clinical Drug Comp
21141600	clopidogrel 1 MG/ML	RxNorm Extension	Clinical Drug Comp
36222254	clopidogrel Oral Product	RxNorm	Clinical Dose Group
21092477	clopidogrel 5 MG/ML	RxNorm Extension	Clinical Drug Comp
21177192	100 ML clopidogrel 1 MG/ML Oral Suspension	RxNorm Extension	Quant Clinical Drug
21047899	1 ML clopidogrel 5 MG/ML Oral Suspension	RxNorm Extension	Quant Clinical Drug
21121870	clopidogrel 5 MG/ML Oral Suspension	RxNorm Extension	Clinical Drug
21063106	clopidogrel 75 MG Oral Tablet [Grepid]	RxNorm Extension	Branded Drug
1322190	clopidogrel 300 MG Oral Tablet [Plavix]	RxNorm	Branded Drug
21121869	clopidogrel 75 MG Injectable Solution	RxNorm Extension	Clinical Drug
21053280	clopidogrel 6 MG Injectable Solution	RxNorm Extension	Clinical Drug
21023810	clopidogrel 4 MG Injectable Solution	RxNorm Extension	Clinical Drug
21106783	1 ML clopidogrel 1 MG/ML Oral Suspension	RxNorm Extension	Quant Clinical Drug
19075601	clopidogrel 75 MG Oral Tablet	RxNorm	Clinical Drug
21102364	clopidogrel 1 MG/ML Oral Suspension	RxNorm Extension	Clinical Drug
40095879	clopidogrel Oral Tablet [Plavix]	RxNorm	Branded Drug Form
40095878	clopidogrel Oral Tablet	RxNorm	Clinical Drug Form
21088717	100 ML clopidogrel 15 MG/ML Oral Suspension	RxNorm Extension	Quant Clinical Drug

Vocabulary classifications improve your efficiency….and your quality!

DeFalco HSORM 2013

Table 3 Identification of related 11 digit NDC codes by drug class and vocabulary

Drug class	Vocabulary	System grouping	Ingredients	Clinical drugs	NDC codes	Unique codes
Opioid	ATC	Opioids	23	1,122	11,765	2
Opioid	ETC.	Analgesics–narcotic	20	1,808	19,106	333
Opioid	NDFRT	Opioid agonists	22	1,813	15,912	1,087
Opioid	VA	Opioid analgesics	24	1,750	17,113	450
NSAID	ATC	Antiinflam and antirheumatic products, non-steroids	52			
NSAID	ETC.	NSAID analgesics	23			
NSAID	NDFRT	NSAID analgesics	23			
NSAID	VA	Nonsalicylate NSAIDs, antirheumatic	24			
Antidiabetic	ATC	Drugs used in diabetes	53			
Antidiabetic	ETC.	Oral antidiabetic agents	19			
Antidiabetic	NDFRT	Insulin receptor agonists	42			
Antidiabetic	VA	Oral hypoglycemic agents	18			
Antidepressant	ATC	Antidepressants	47			
Antidepressant	ETC.	Antidepressants	29			
Antidepressant	NDFRT	Serotonin uptake inhibitors, norepinephrine uptake inhibitors, dopamine uptake inhibitors	40			
Antidepressant	VA	Antidepressants	29			

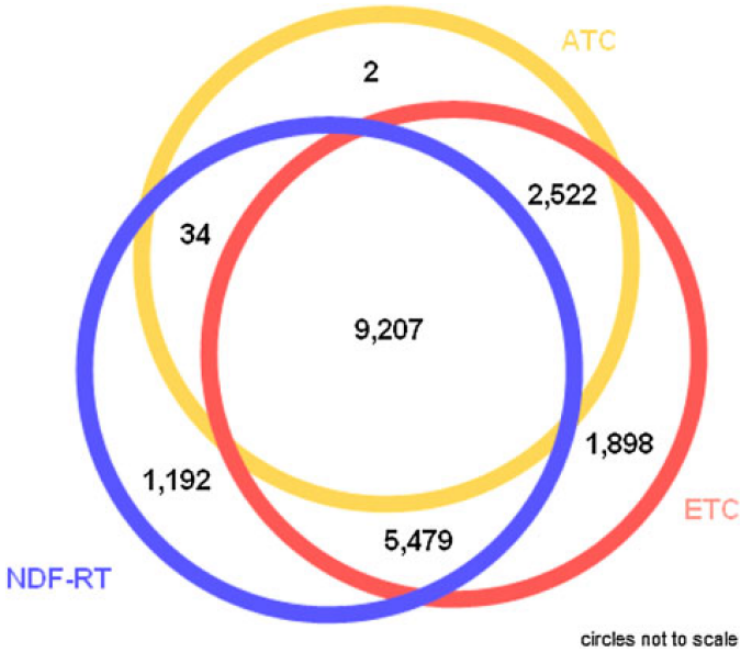


Fig. 1 Overlap in coverage of 'opioid' NDC drug codes by classification system

Find members of Drug Classes

4. Check Ingredient Descendants of Drug Class Anticoagulants

```
SELECT max_levels_of_separation, concept.*
FROM concept_ancestor
JOIN concept ON descendant_concept_id = concept_id
WHERE ancestor_concept_id = 21600961 /* ATC Antithromboic Agent */
    AND concept_class_id = 'Ingredient'
ORDER BY max_levels_of_separation
```

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id
46275677	cangrelor	Drug	RxNorm	Ingredient
45892847	edoxaban	Drug	RxNorm	Ingredient
1322184	clopidogrel	Drug	RxNorm	Ingredient
44818499	vorapaxar	Drug	RxNorm	Ingredient
43013024	apixaban	Drug	RxNorm	Ingredient
42898933	defibrotide	Drug	RxNorm	Ingredient
42801108	Protein C	Drug	RxNorm	Ingredient
40241331	rivaroxaban	Drug	RxNorm	Ingredient
1310149	Warfarin	Drug	RxNorm	Ingredient
40241186	Ticagrelor	Drug	RxNorm	Ingredient
40228152	dabigatran etexilate	Drug	RxNorm	Ingredient
40163718	prasugrel	Drug	RxNorm	Ingredient
35604848	selexipag	Drug	RxNorm	Ingredient
19136187	Streptokinase	Drug	RxNorm	Ingredient

If we try to speak the same language, will there be loss in translation?

Journal of Biomedical Informatics 45 (2012) 689–696

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Evaluation of alternative standardized terminologies for medical conditions within a network of observational healthcare databases ☆

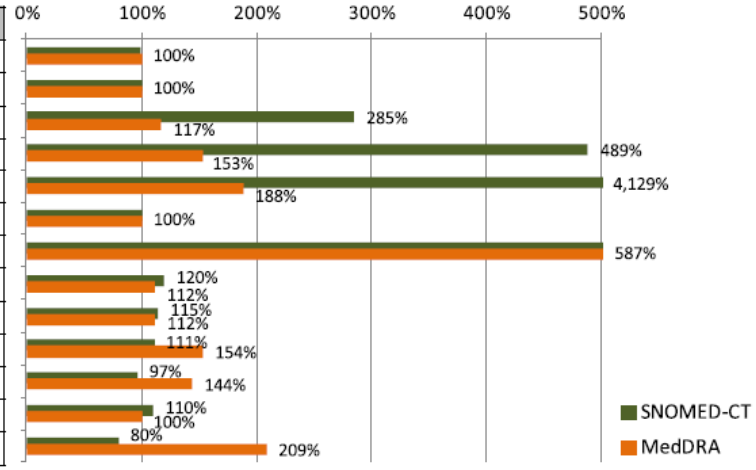
Christian Reich [a,*], Patrick B. Ryan [a,b,1], Paul E. Stang [a,b,1], Mitra Rocca [c,2]

[a] Observational Medical Outcomes Partnership, Foundation for the National Institutes of Health, 9650 Rockville Pike, Bethesda, MD 20814, USA
[b] Janssen Research & Development, LLC, 1125 Trenton-Harbourton Road, PO Box 200, MS K304, Titusville, NJ 08560, USA
[c] Office of Translational Sciences, Center for Drug Evaluation and Research (CDER), US Food and Drug Administration, 10903 New Hampshire Ave., Bldg. 21, Rm. 4608, Silver Spring, MD 20933, USA

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

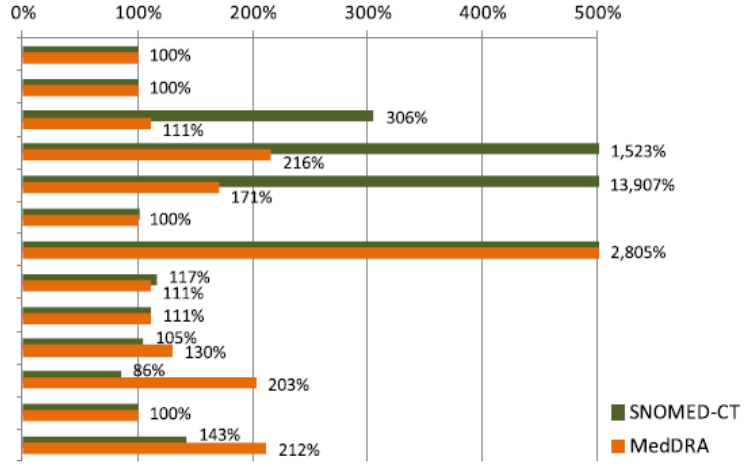
Cohort size of HOI in MSLR for different terminologies

Condition	#
Angioedema	1
Aplastic Anemia	1
Acute Liver Failure	1
	2
	7
Acute Renal Failure	1
	4
Bleeding	1
	3
Hip Fracture	1
Acute myocardial Infarction	1
	2
Upper GI Ulcer Hospitalization	1



Cohort size of HOI in GE for different terminologies

Condition	#
Angioedema	1
Aplastic Anemia	1
Acute Liver Failure	1
	2
	7
Acute Renal Failure	1
	4
Bleeding	1
	3
Hip Fracture	1
Acute myocardial Infarction	1
	2
Upper GI Ulcer Hospitalization	1



…but in practice, running an estimation analysis using source vs. standard vocabulary yields similar results

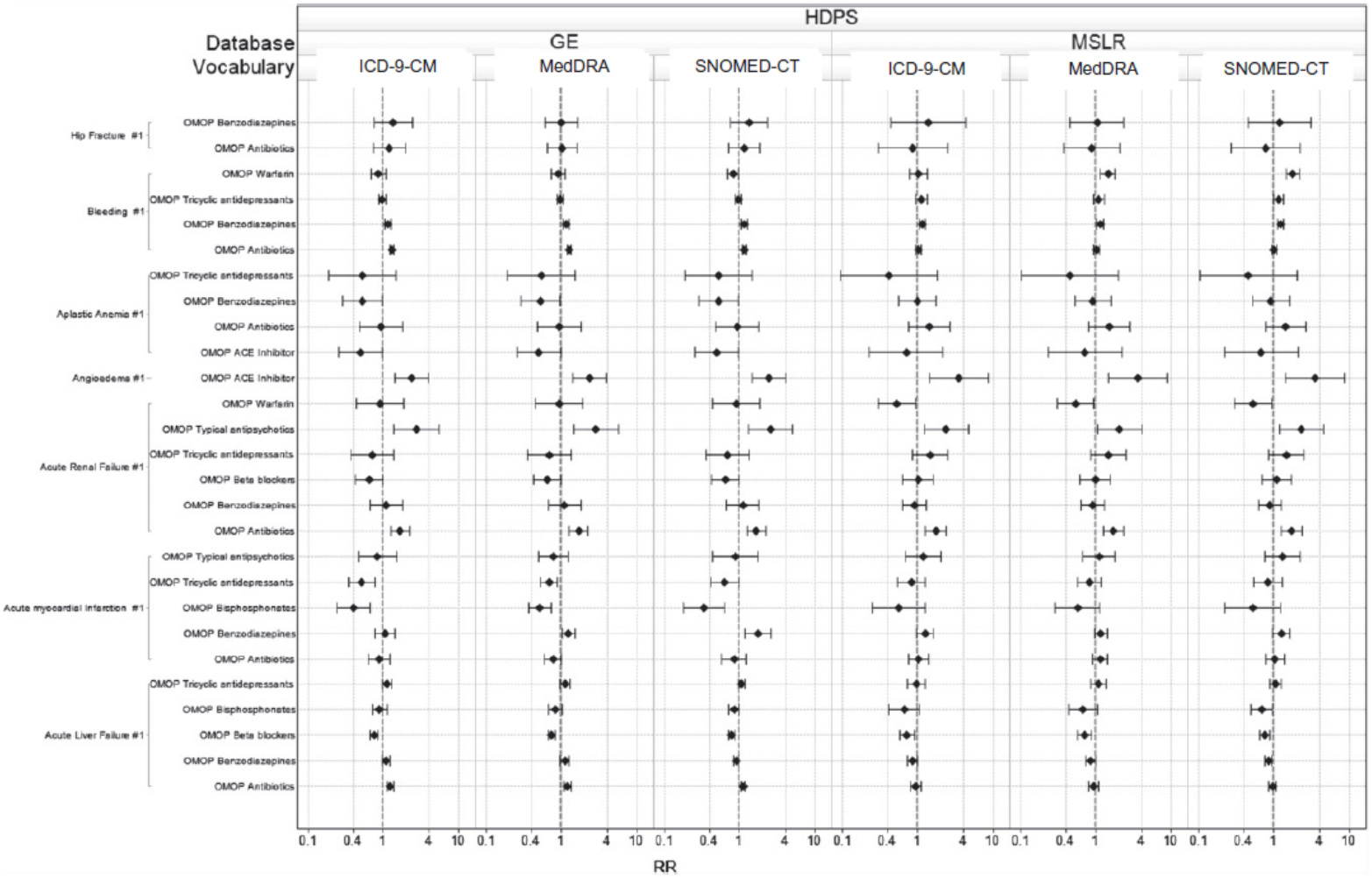


Fig. 3. Effect estimates and 95% confidence intervals for incident user design applied to MSLR and GE using ICD-9-CM, SNOMED-CT, and MedDRA as standard terminologies. Each dot represents the estimate of the effect of an individual HOI-drug combination (on the X-axis).

Defining cohorts

Defining ‘phenotype’

Journal of the American Medical Informatics Association, 0(0), 2017, 1–6
doi: 10.1093/jamia/ocx110
Perspective

Perspective

High-fidelity phenotyping: richness and freedom from bias

George Hripcsak[1] and David J Albers[1]

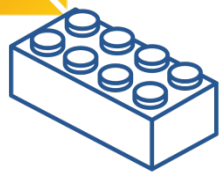
- A phenotype is a specification of an observable, potentially changing state of an organism (as distinguished from the genotype, derived from genetic makeup).
- The term phenotype can be applied to patient characteristics inferred from electronic health record (EHR) data.
- The goal is to draw conclusions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

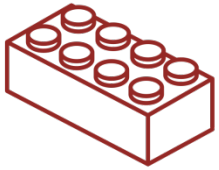
Rule-Based Phenotyping

Probabilistic Phenotyping

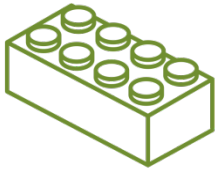
Data are Like Lego Bricks for Phenotypng



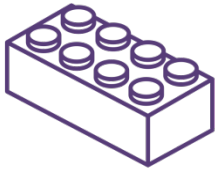
Conditions



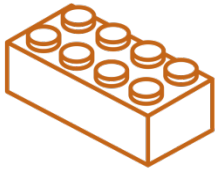
Drugs



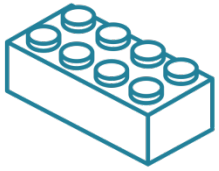
Procedures



Measurements



Observations



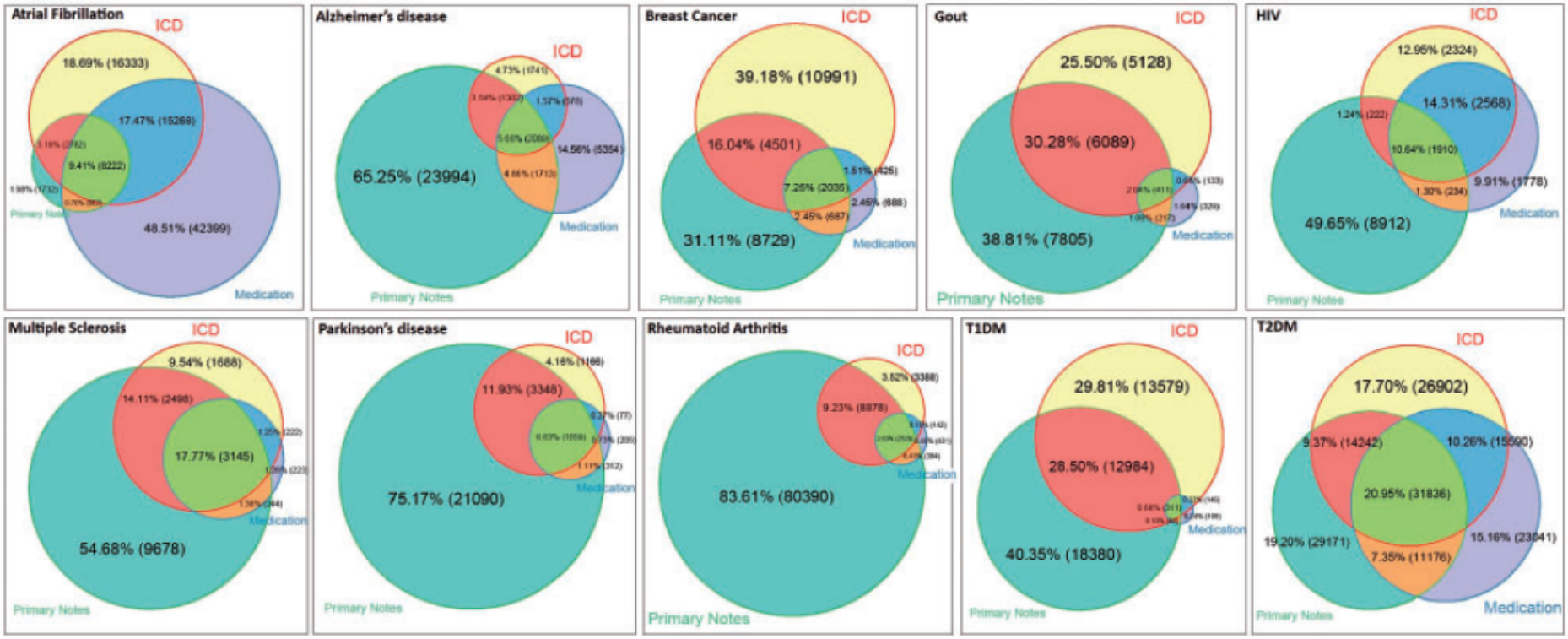
Visits

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 Wei[1], Pedro L Teixeira[1], Huan Mo[1], 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
1^0 dx of any CHF	19 339	54.8	1.9	86.0	2.2
1^0 dx of 428	16 724	47.6	1.7	86.3	2.5
1^0 dx of any CHF and BNP >500 pg/mL	11 298	33.5	1.3	90.0	2.1
1^0 dx of 428 and BNP >500 pg/mL	9662	28.8	1.1	90.4	2.4
1^0 dx of 428 and BNP >500 pg/mL and echocardiogram	5678	16.2	0.8	86.6	3.5
1^0 dx of any CHF or BNP >500 pg/mL	29 587	71.4	2.1	73.3	2.2
1^0 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'

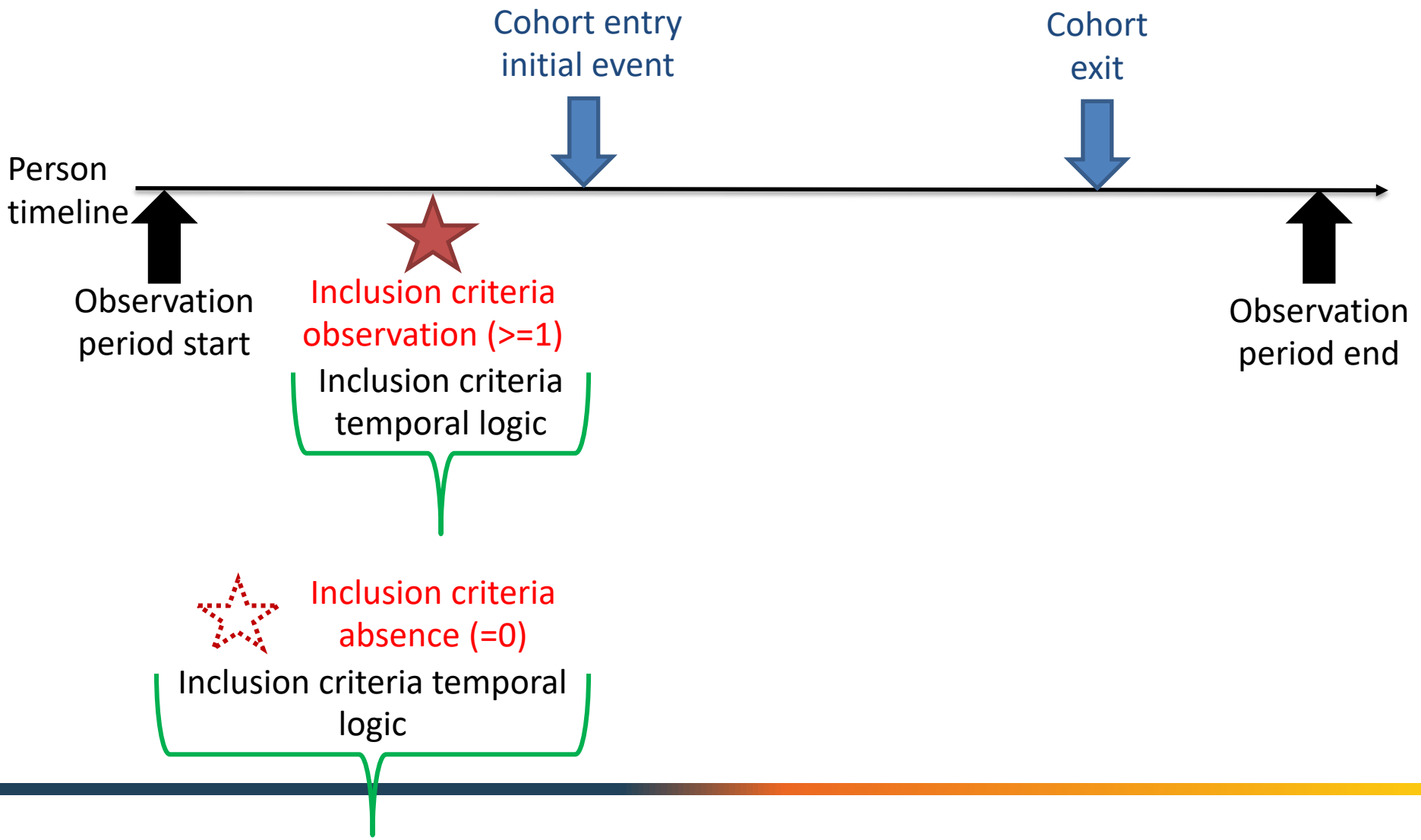
Cohort = a set of persons who satisfy one or more inclusion criteria for a duration of time

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

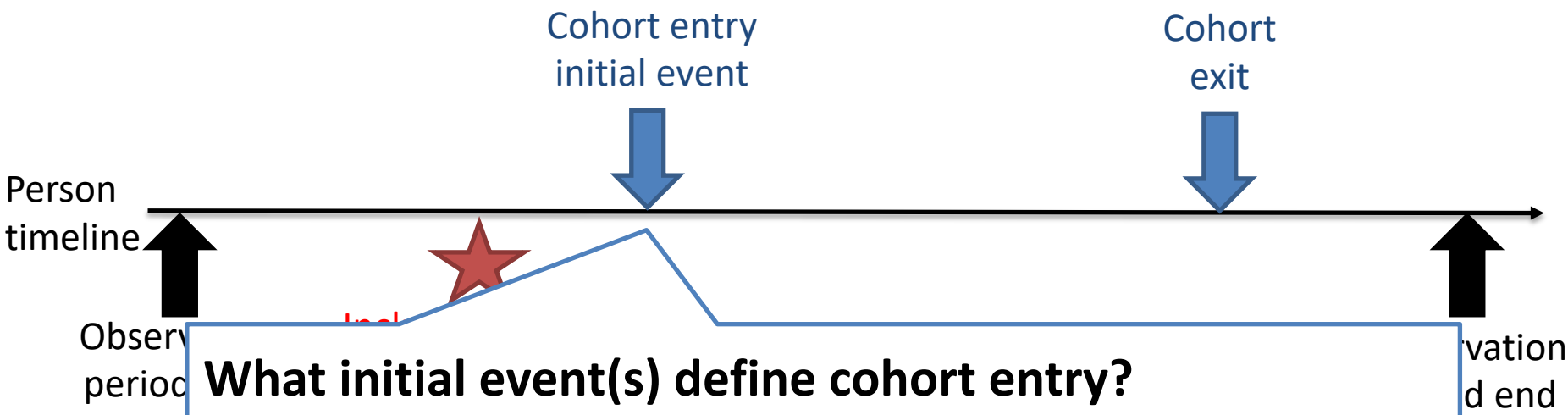
Dissecting the anatomy of a cohort definition



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?

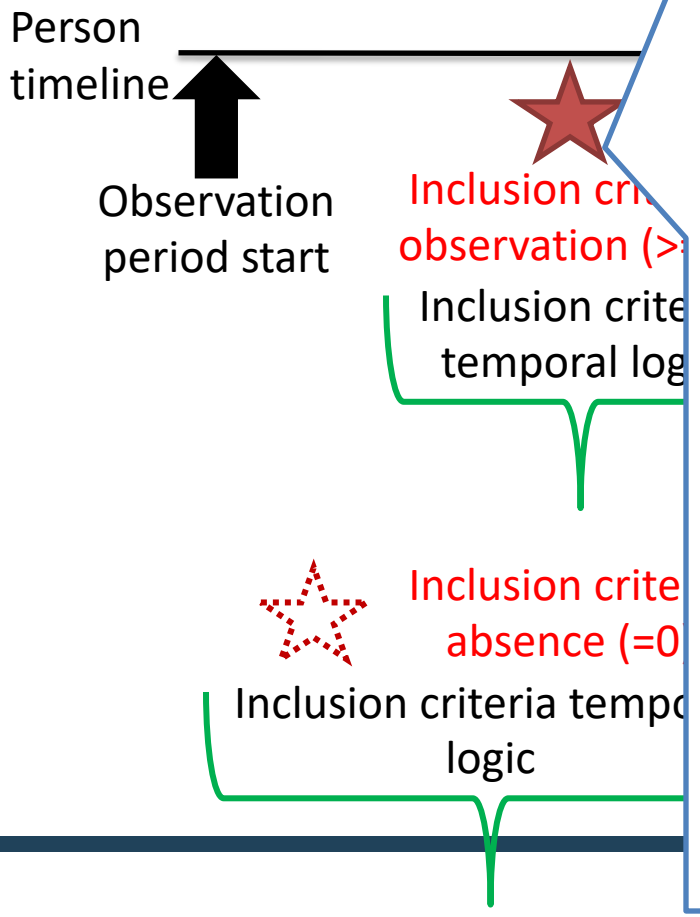
Dissecting the anatomy of a cohort definition



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

Dissecting the anatomy of a cohort definition



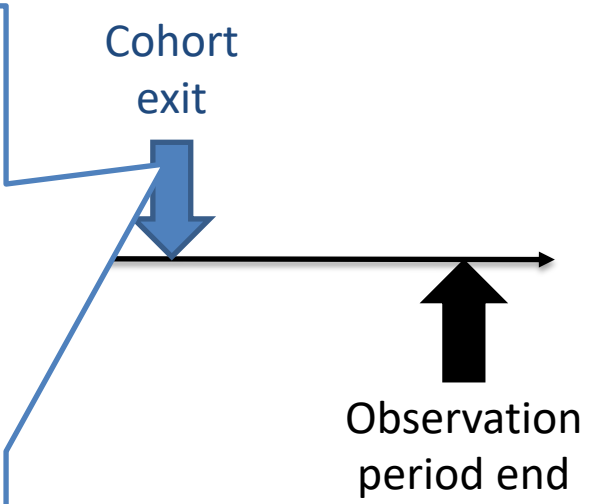
What inclusion criteria are applied to the initial events?

- 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

Implementing cohort definitions in ATLAS

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

Explore a data source for available data elements

Graham et al. description of the outcomes

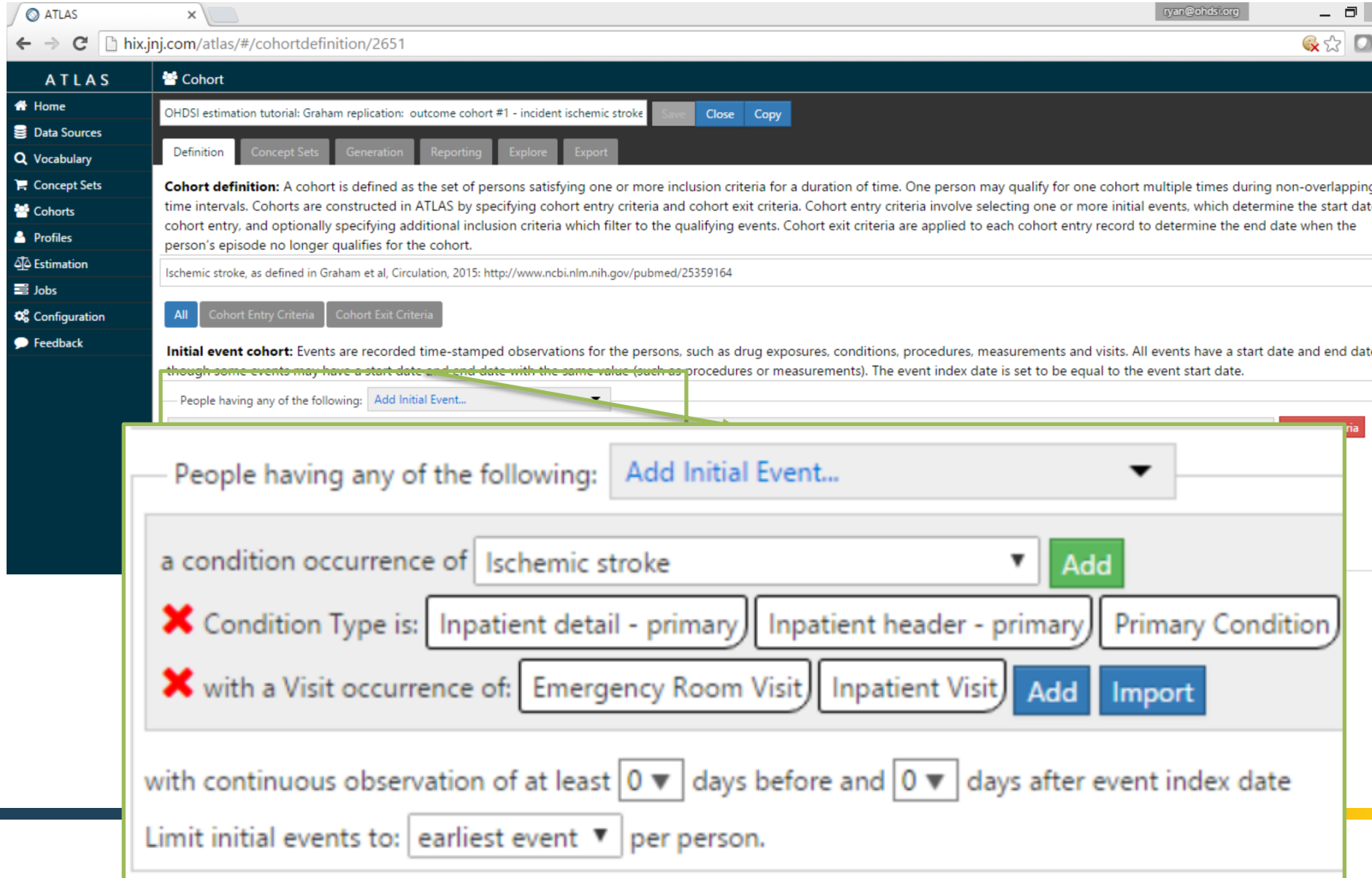
Study Outcomes

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

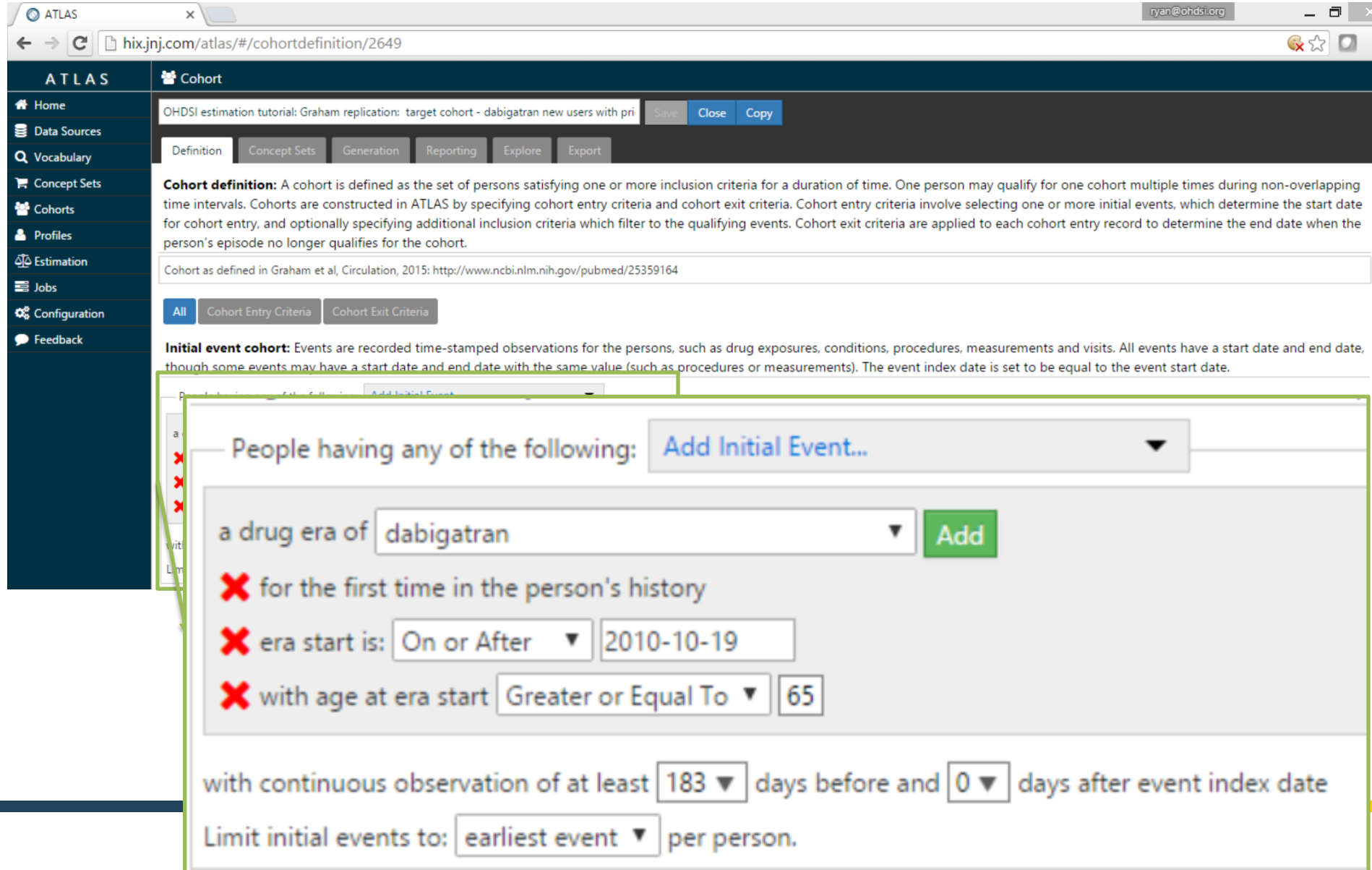
Graham et al. replication: Designing the outcome cohort in ATLAS



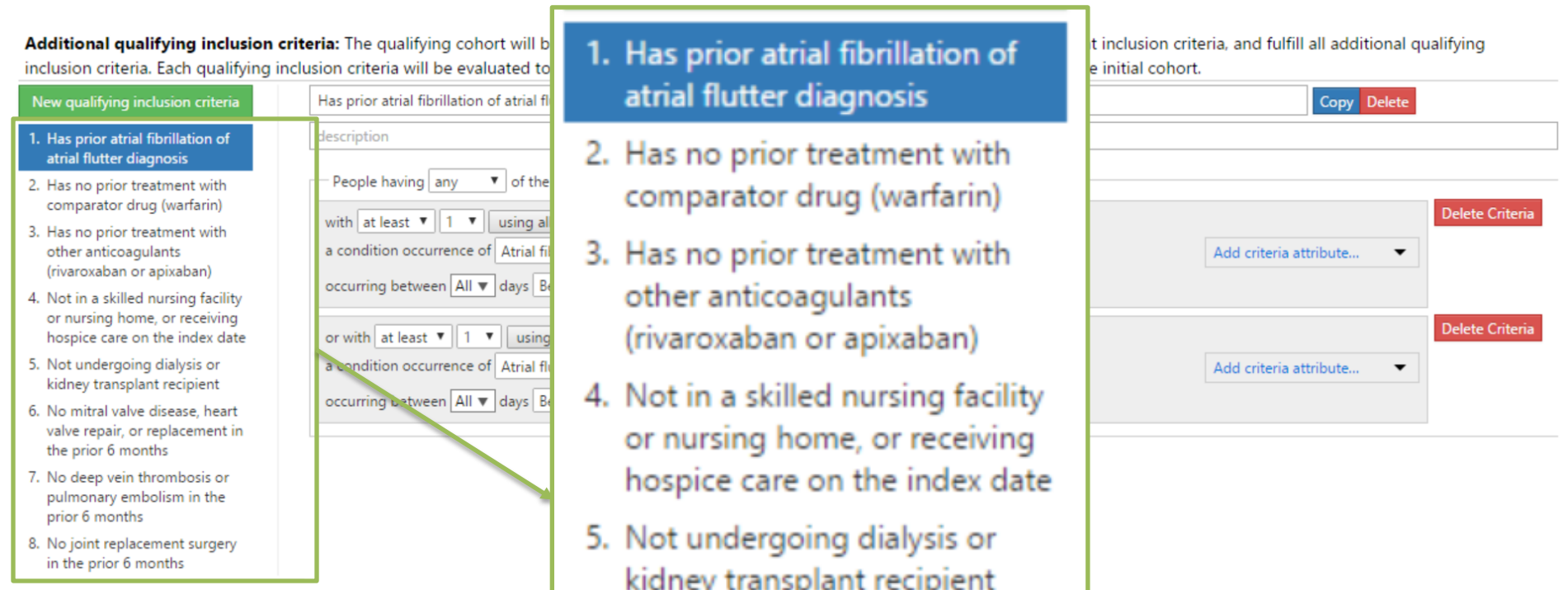
Graham et al. description of the cohort(s)

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.

Graham et al. replication: Designing the target cohort in ATLAS



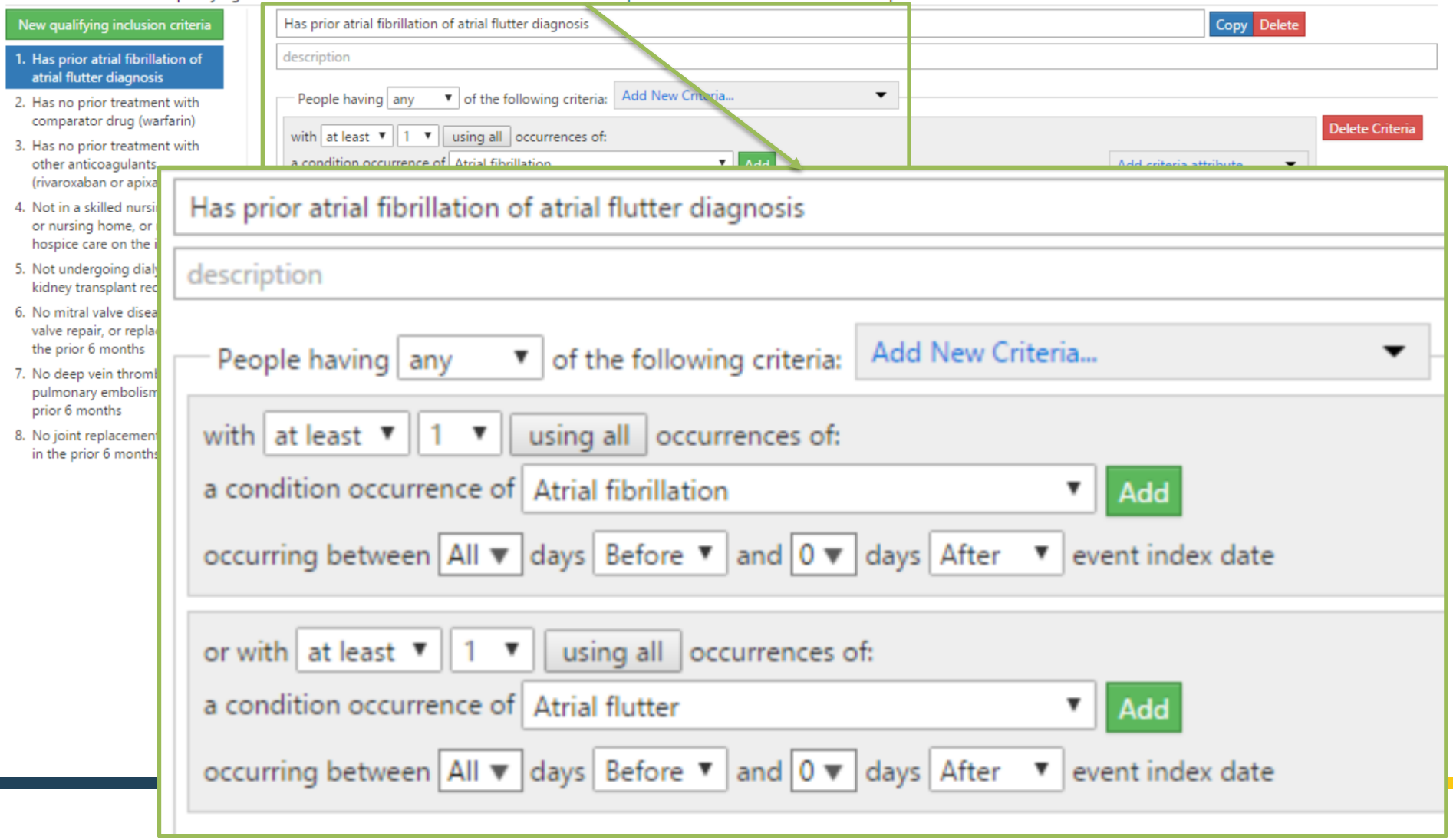
Graham et al. replication: Designing the target cohort in ATLAS



1. Has prior atrial fibrillation of atrial flutter diagnosis
2. Has no prior treatment with comparator drug (warfarin)
3. Has no prior treatment with other anticoagulants (rivaroxaban or apixaban)
4. Not in a skilled nursing facility or nursing home, or receiving hospice care on the index date
5. Not undergoing dialysis or kidney transplant recipient
6. No mitral valve disease, heart valve repair, or replacement in the prior 6 months
7. No deep vein thrombosis or pulmonary embolism in the prior 6 months
8. No joint replacement surgery in the prior 6 months

Graham et al. replication: Designing the target cohort in ATLAS

Additional qualifying inclusion criteria: The qualifying cohort will be defined as all persons who have an initial event, satisfy the initial event inclusion criteria, and fulfill all additional qualifying inclusion criteria. Each qualifying inclusion criteria will be evaluated to determine the impact of the criteria on the attrition of persons from the initial cohort.



Graham et al. replication: Designing the target cohort in ATLAS

Cohort Exit Criteria

Cohort exit criteria based on the end of an era of persistent exposure to any drug within a defined concept set:

Specify a concept set that contains one or more drugs. A drug era will be derived from all drug exposure events for any of the drugs within the concept set, using the specified persistence window as a maximum allowable gap in days between successive exposure events and adding a specified surveillance window to the final exposure event. If no exposure event end date is provided, then an exposure event end date is inferred to be event start date + days supply in cases when days supply is available or event start date + 1 day otherwise. This cohort exit criteria assures that the cohort end date will be no greater than the drug era end date.

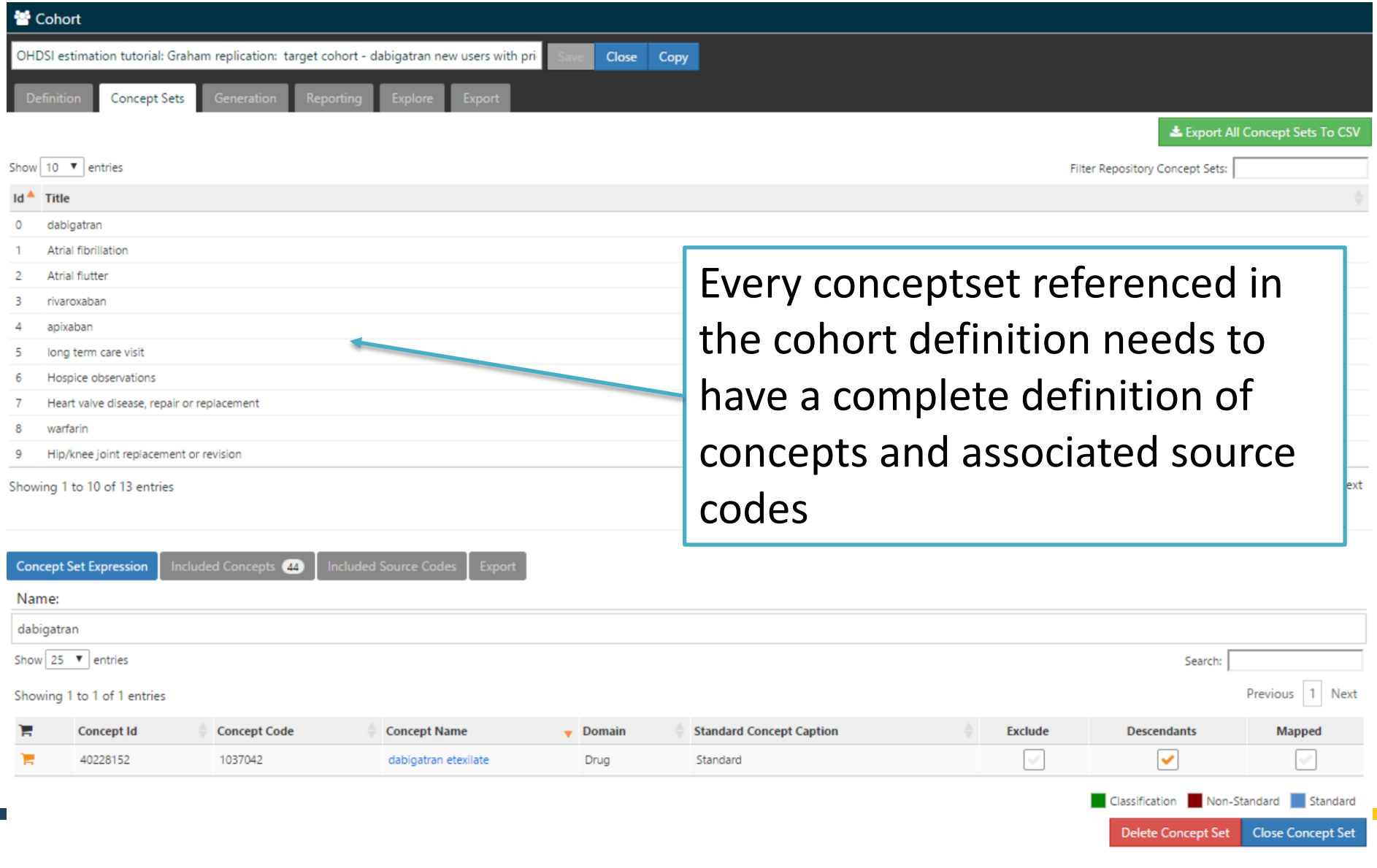
Concept set containing the drug(s) of interest: dabigatran ▼ Add

- Persistence window: allow for a maximum of 3 ▼ days between exposure records when inferring the era of persistence exposure
- Surveillance window: add 0 ▼ days to the end of the era of persistence exposure as an additional period of surveillance prior to cohort exit.

Concept set containing the drug(s) of interest: dabigatran ▼ Add

- Persistence window: allow for a maximum of 3 ▼ days between exposure records when inferring the era of persistence exposure
- Surveillance window: add 0 ▼ days to the end of the era of persistence exposure as an additional period of surveillance prior to cohort exit.

Graham et al. replication: Designing the target cohort in ATLAS



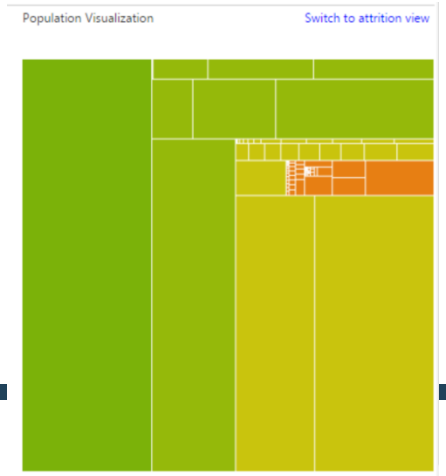
Every conceptset referenced in the cohort definition needs to have a complete definition of concepts and associated source codes

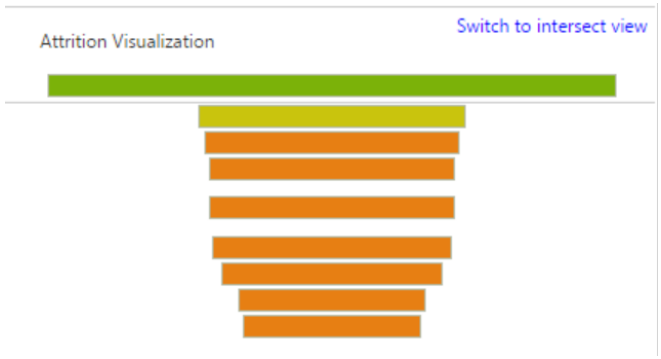
Graham et al. replication: Evaluating the impact of inclusion criteria on the cohort in ATLAS

Inclusion Report for Truven MDCR

	Match Rate	Matches	Total
Summary Statistics:	31.52%	52,400	166,243

	Inclusion Rule	N	% Satisfied	% To-Gain
1.	Has prior atrial fibrillation of atrial flutter diagnosis	78,371	47.14%	16.40%
2.	Has no prior treatment with comparator drug (dabigatran)	162,601	97.81%	1.44%
3.	Has no prior treatment with other anticoagulants (rivaroxaban or apixaban)	161,768	97.31%	1.26%
4.	Not in a skilled nursing facility or nursing home, or receiving hospice care on the index date	166,149	99.94%	0.01%
5.	Not undergoing dialysis or kidney transplant recipient	163,463	98.33%	0.65%
6.	No mitral valve disease, heart valve repair, or replacement in the prior 6 months	157,221	94.57%	2.91%
7.	No deep vein thrombosis or pulmonary embolism in the prior 6 months	118,058	71.02%	5.56%
8.	No joint replacement surgery in the prior 6 months	138,630	83.39%	1.44%



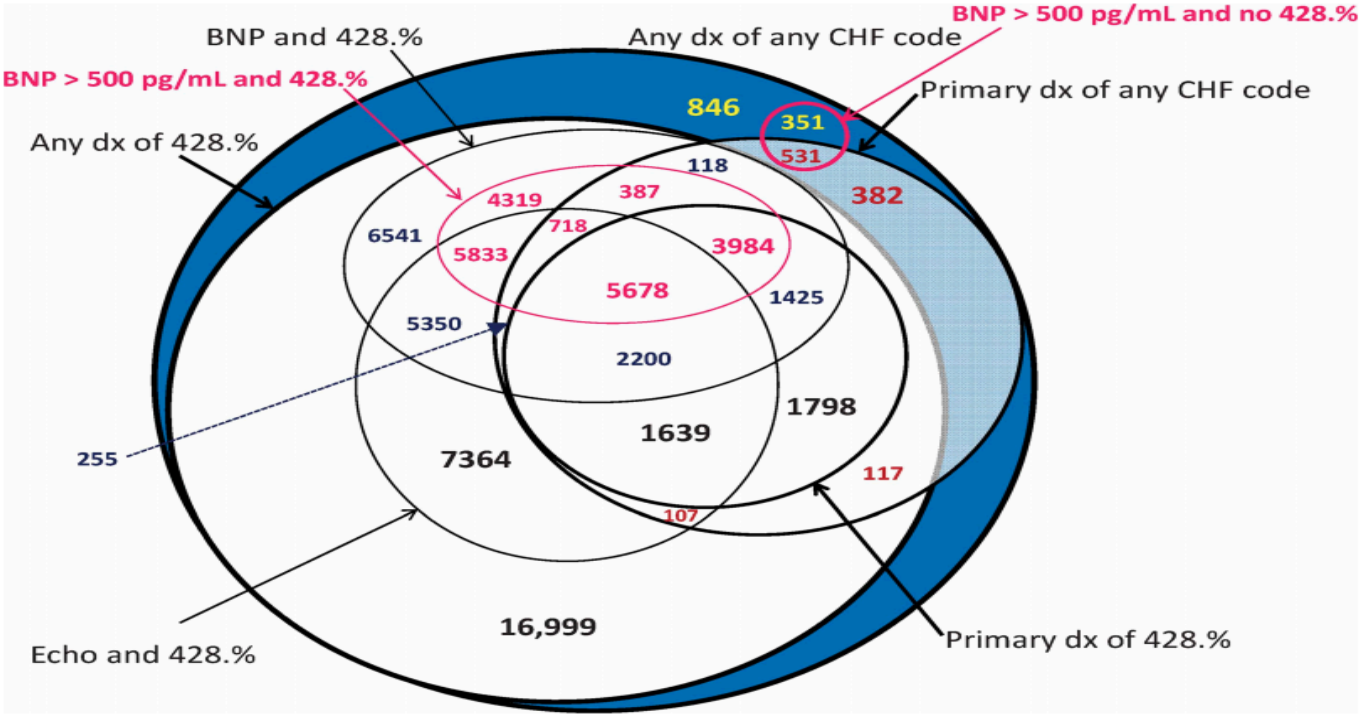


Evaluating cohort definitions

Evaluating cohort definitions

- How do we know if our cohort definition is any “good”?
- What is our goal for a cohort definition’s performance for a given use case?
- How do we know if our cohort definition is generalizable across sites?

Example: Evaluating CHF definitions



Rosenman et al. Database queries for hospitalizations for acute congestive heart failure: flexible methods and validation based on set theory. *J Am Med Inform Assoc*. 2014 Mar-Apr;21(2):345-52.

Example: Evaluating CHF definitions

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
1^0 dx of any CHF	19 339	54.8	1.9	86.0	2.2
1^0 dx of 428	16 724	47.6	1.7	86.3	2.5
1^0 dx of any CHF and BNP >500 pg/mL	11 298	33.5	1.3	90.0	2.1
1^0 dx of 428 and BNP >500 pg/mL	9662	28.8	1.1	90.4	2.4
1^0 dx of 428 and BNP >500 pg/mL and echocardiogram	5678	16.2	0.8	86.6	3.5
1^0 dx of any CHF or BNP >500 pg/mL	29 587	71.4	2.1	73.3	2.2
1^0 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.

Ground Truth?

- To measure performance we need an outcome such as ‘case’ and ‘not a case’
- This determination is typically based on expert review of available data (e.g., sometimes will have notes etc that are not part of definition)
- The review process may include some heuristic guidance to ensure consistency amongst reviewers + Cohen’s Kappa
- Some newer research into automated ways to assess true cases (e.g., cohort characteristics)

Performance Metrics

- PPV is currently the primary metric obtained through manual review
- Sensitivity is sometimes determined when there are sufficient resources or when the incidence rate is reasonably high

Graham et al's discussion of outcome 'validation'

The codes defining ischemic stroke have a positive predictive value (PPV) of 88% to 95%.[18–20] Major bleeding was defined as a fatal bleeding event, a hospitalized bleeding event requiring transfusion,
(ie, intracra
retroperiton
Intracranial
atraumatic
for hemorr
validated. V
a bleeding e
related. The
86% to 88%
a PPV of 8
and 97% in

Rewriting to state our knowledge about data quality:

"Somewhere between 1-in-10 and 1-in-20 patients who have one of the diagnosis codes for ischemic stroke DO NOT actually have ischemic stroke."

"We DO NOT know how many people who don't have the stroke diagnosis codes actually DO have ischemic strokes (e.g. missing data, miscoding, censoring – death before heath service utilization), or whether these false negatives represent a differential bias."

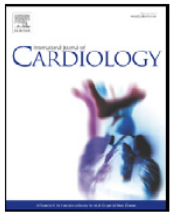
Validating Algorithms

Many research studies have attempted to validate algorithms

Contents lists available at ScienceDirect

International Journal of Cardiology

journal homepage: www.elsevier.com/locate/ijcard



Review

Use of electronic health records to ascertain, validate and phenotype acute myocardial infarction: A systematic review and recommendations

Bruna Rubbo [a,*], Natalie K. Fitzpatrick [a], Spiros Denaxas [a], Marina Daskalopoulou [b], Ning Yu [a], Riyaz S. Patel [a,c], UK Biobank Follow-up and Outcomes Working Group, Harry Hemingway [a]

- Examined 33 studies
- Found significant heterogeneity in algorithms used, validation methods, and results

Validating an Algorithm

		Truth	
		Positive	**Negative**
Test	**Positive**	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Test – Comes from the algorithm/cohort definition

Truth – Some form of "gold standard" reference

Ex.: True Positive (TP) – Test and Truth agree Positive

For a complete validation of the algorithm we need:

1) Sensitivity: TP / (TP + FN)
2) Specificity: TN / (TN + FP)
3) Positive Predictive Value: TP / (TP + FP)

Evaluating Performance of Algorithm - Examples

Abstract

Purpose—To validate an algorithm based upon International Classification of Diseases, 9th revision, Clinical Modification (ICD-9-CM) codes for acute myocardial infarction (AMI) documented within the Mini-Sentinel Distributed Database (MSDD).

Methods—**Using an ICD-9-CM-based algorithm (hospitalized patients with 410.x0 or 410.x1 in primary position),** we identified a random sample of potential cases of AMI in 2009 from 4 Data Partners participating in the Mini-Sentinel Program. **Cardiologist reviewers used information abstracted from hospital records to assess the likelihood of an AMI diagnosis based on criteria from the joint European Society of Cardiology and American College of Cardiology Global Task Force.** Positive predictive values (PPVs) of the ICD-9-based algorithm were calculated.

Results—Of the 153 potential cases of AMI identified, hospital records for 143 (93%) were retrieved and abstracted. **Overall, the PPV was 86.0% (95% confidence interval; 79.2%, 91.2%).** PPVs ranged from 76.3% to 94.3% across the 4 Data Partners.

Conclusions—The overall PPV of potential AMI cases, as identified using an ICD-9-CM-based algorithm, may be acceptable for safety surveillance; however, PPVs do vary across Data Partners. This validation effort provides a contemporary estimate of the reliability of this algorithm for use in future surveillance efforts conducted using the FDA's MSDD.

Evaluating Performance of Algorithm - Examples

PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2009; **18**: 1064–1071
Published online 28 August 2009 in Wiley InterScience (www.interscience.wiley.com) **DOI**: 10.1002/pds.1821

ORIGINAL REPORT

SUMMARY

Purpose Studies of non-steroidal anti-inflammatory drugs (NSAIDs) and cardiovascular events using administrative data require identification of incident acute myocardial infarctions (AMIs) and information on whether confounders differ by NSAID status.

Methods **We identified patients with a first AMI hospitalization from Tennessee Medicaid files as those with primary ICD-9 discharge diagnosis 410.x and hospitalization stay of>2 calendar days.** Eligible persons were non-institutionalized, aged 50–84 years between 1999–2004, had continuous enrollment and no AMI, stroke, or non-cardiovascular serious medical illness in the prior year. Of 5524 patients with a potential first AMI, a systematic sample (n¼350) was selected for review. **Using defined criteria, we classified events using chest pain history, EKG, and cardiac enzymes, and calculated the positive predictive value (PPV) for definite or probable AMI.**

and no AMI, respectively. **PPV for any definite or probable AMI was 92.8% (95% CI 89.6–95.2); for an AMI without an event in the past year 91.7% (95% CI 88.3–94.2), and for an incident AMI was 72.7% (95% CI 67.7–77.2).** Age-adjusted prevalence of current smoking (46.4% vs. 39.1%, p¼0.35) and aspirin use (36.9% vs. 35.9%, p¼0.90) was similar among NSAID users and non-users

Conclusions ICD-9 code 410.x had high predictive value for identifying AMI. Among those with AMI, smoking and aspirin use was similar in NSAID exposure groups, suggesting these factors will not confound the relationship between NSAIDs and cardiovascular outcomes.

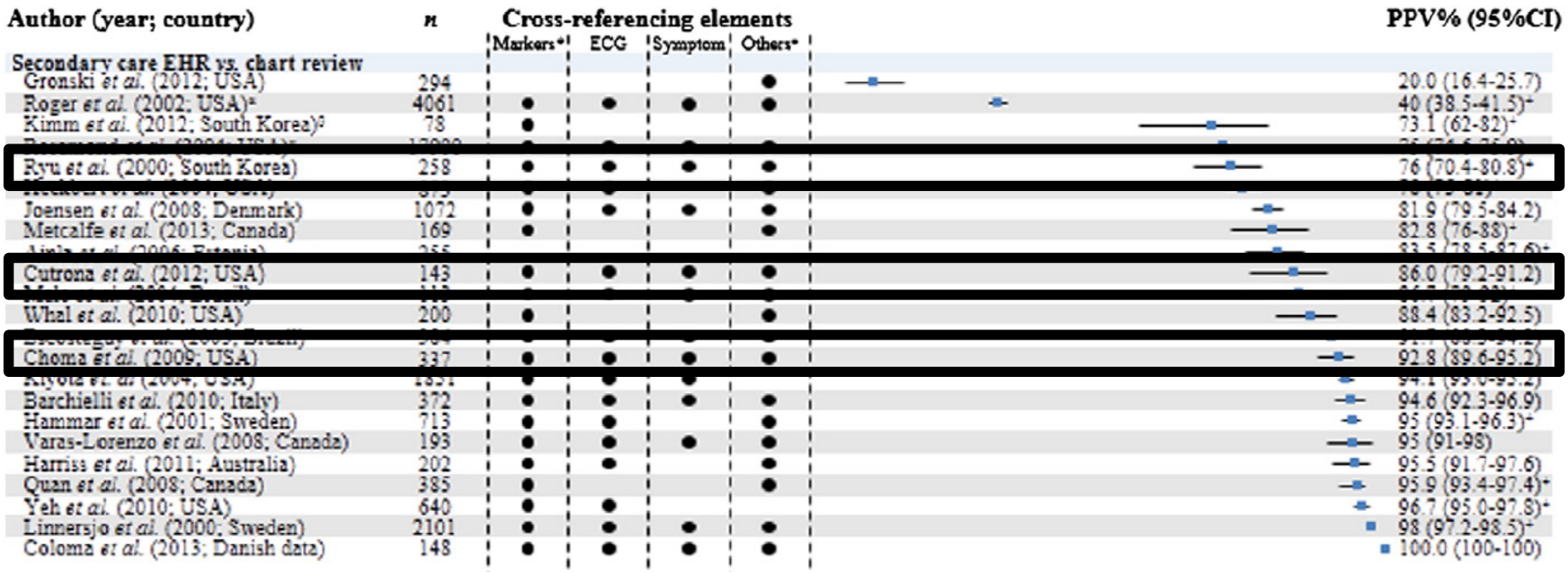
Evaluating Performance of Algorithm - Examples

Yonsei Medical Journal
Vol. 41, No. 5, pp. 570–576, 2000

Abstract

We attempted to assess the accuracy of the International Classification of Diseases (ICD) codes for myocardial infarction (MI) in medical insurance claims, and to investigate the reasons for any inaccuracy. This study was designed as a preliminary study to establish a surveillance system for cardiovascular diseases in Korea. A sample of 258 male patients who were diagnosed with MI from 1993 to 1997 was selected from the Korea Medical Insurance Corporation cohort (KMIC cohort: 183,461 people). The registered medical record administrators were trained in the survey technique, and gathered data by investigating the medical records of the study subjects from March 1999 to May 1999. **The definition of MI for this study included symptoms pursuant to the diagnostic criteria of chest pain, electrocardiogram (ECG) findings, cardiac enzyme and results of coronary angiography or nuclear scan.** We asked the record administrators for the reasons of incorrectness for cases where the final diagnosis was 'not MI'. **The accuracy rate of the ICD codes for MI in medical insurance claims was 76.0% (196 cases) of the study sample,** and 3.9% (ten cases) of the medical records were not available due to hospital closures, non-computerization or missing information. Nineteen cases (7.4%) were classified as insufficient due to insufficient records of chest pain, ECG findings, or cardiac enzymes. The major reason of inaccuracy in the disease code for MI in medical insurance claims was 'to meet the review criteria of medical insurance benefits (45.5%)'. The department responsible for the inaccuracy was the department of inspection for medical insurance benefit of the hospitals.

Evaluating Performance of Algorithm



Evaluating Performance of Algorithm

- Conclusion – for MI → no "gold standard" algorithm available
- Process is very costly and time consuming
- Small sample sizes → wide variation in estimates with wide confidence intervals

- In 33 studies "validating" algorithms, all reported PPV but:
  - Only 11 reported sensitivity
  - Only 5 reported specificity
  - **Is this really validation?**

The Value of Positive Predictive Value

- PPV is almost always reported in validation studies – easiest to assess
- Sensitivity and Specificity much less frequently reported
  - High cost and time to evaluate
- BUT – sensitivity and specificity are the actual characteristics of the test
  - PPV is a function of sensitivity, specificity and **prevalence** of Heath Outcome of Interest (HOI)

PPV Example – 1 Test, 2 Populations

Test Characteristics:
Sensitivity = 75%
Specificity = 99.9%

Population = 10,000

Prevalence = 1%		Truth	
		Positive	Negative
Test	Positive	75	10
	Negative	25	9890
	Total	100	9900

PPV =
75 / (75 + 10) =
88%

Prevalence = 5%		Truth	
		Positive	Negative
Test	Positive	375	10
	Negative	125	9490
		500	9500

PPV =
375 / (375 + 10) =
97%

PPV Example – 1 Population, 2 Tests

PPV = 90%

Population = 10,000

Prevalence = 5%		Truth	
		Positive	Negative
Test #1	Positive	90	10
	Negative	410	9490
	Total	500	9500

PPV = 90/(90+10) = 90%

Sens = 90/500 = 18%

Spec = 9490/9500 = 99.9%

Prevalence = 5%		Truth	
		Positive	Negative
Test #2	Positive	360	40
	Negative	140	9460
		500	9500

PPV = 360/(360+40) = 90%

Sens = 360/500 = 72%

Spec = 9460/9500 = 99.6%

Living with Algorithms

- Algorithms are used in most research with observational data
- Many ways to define an algorithm for any health outcome
- Each definition will have its own performance characteristics
  - Need to validate the algorithm to understand these characteristics
- Validation of an algorithm to be used in an observational dataset through chart review is likely not possible
  - Costly
  - Time consuming
  - Data is usually not available

Areas to improve cohort validation

- Cohort characterization
- Electronic patient profile adjudication
- Probabilistic phenotype evaluation

Evaluating cohorts: characterization

Let’s use what we’ve learned to help support some patient-level prediction problems

Questions?

Thanks for joining
the journey!

ryan@ohdsi.org