OHDSI China Tutorial 2018

Mui van Zandt, Christian Reich, Hua Xu

王琼，董楠，吴承凯，张亮军，罗敏
王理，王志，唐灵逸，尚勇，郝天永

OMOP 通用数据模型及标准化术语集

本次教程结束后，你将了解…

1. 什么是 OMOP, OHDSI?
2. 标准化术语集是如何运作的?
3. 怎样找到代码和概念?
4. 如何浏览层级结构?
5. 什么是 OMOP CDM?
6. 如何使用 OMOP CDM

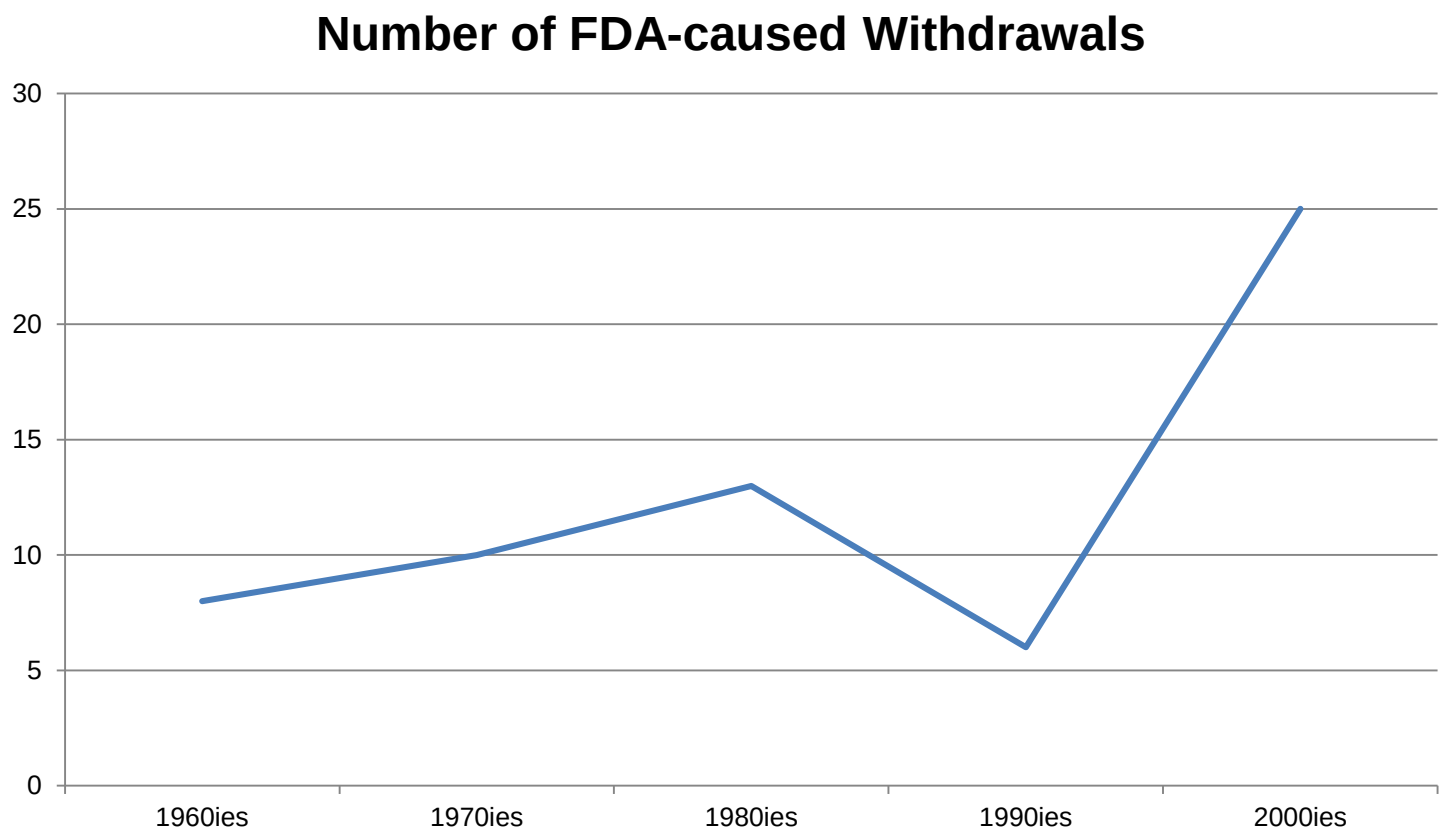
Agenda

Section	Time	Item(s)
Vocabulary	14:00 - 16:00 (1.5 hour)	OHDSI Community Basic Relationship Ancestors & Descendants How does it work for Drugs
Break	16:00 – 16:15	Break
CDM	16:15 – 17:15 (1 hour)	In depth discussion of model Era discussion Real World Scenario
ETL	17:15 – 18:00 (45 min)	ETL Process ETL Tools Potential Issues

健康观测数据科学和信息联盟(OHDSI)计划和志向

Observational Medical Outcomes Partnership

FDA Regulatory Action over Time



Observational Medical Outcomes Partnership

FDAAA calls for establishing Risk Identification and Analysis System

SEC. 905. ACTIVE POSTMARKET RISK IDENTIFICATION AND ANALYSIS.

(a) IN GENERAL.—Subsection (k) of section 505 of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. 355) is amended by adding at the end the following:

"(3) ACTIVE POSTMARKET RISK IDENTIFICATION.—

"(A) DEFINITION.—In this paragraph, the term 'data' refers to information with respect to a drug approved under this section or under section 351 of the Public Health Service Act, including claims data, patient survey data, standardized analytic files that allow for the pooling and analysis of data from disparate data environments, and any other data deemed appropriate by the Secretary.

"(B) DEVELOPMENT OF POSTMARKET RISK IDENTIFICATION AND ANALYSIS METHODS.—The Secretary shall, not later than 2 years after the date of the enactment of the Food and Drug Administration Amendments Act of 2007, in collaboration with public, academic, and private entities—

"(i) develop methods to obtain access to disparate data sources including the data sources specified in subparagraph (C);

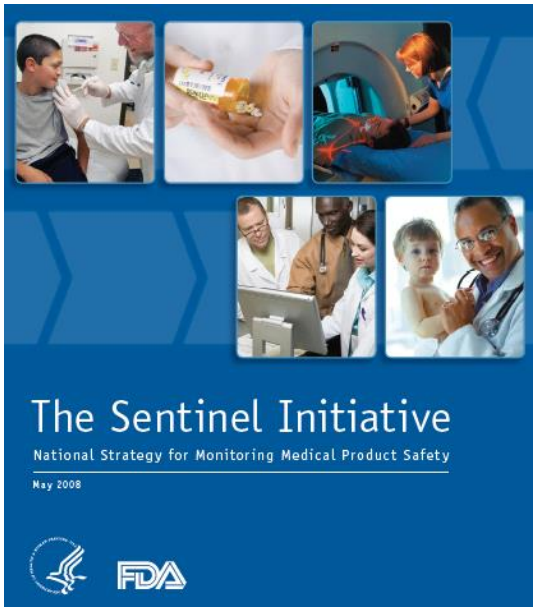
"(ii) develop validated methods for the establishment of a postmarket risk identification and analysis system to link and analyze safety data from multiple sources, with the goals of including, in aggregate—

"(I) at least 25,000,000 patients by July 1, 2010; and

"(II) at least 100,000,000 patients by July 1, 2012; and

"(iii) convene a committee of experts, including individuals who are recognized in the field of protecting data privacy and security, to make recommendations to the Secretary on the development of tools and methods for the ethical and scientific uses for, and communication of, postmarketing data specified under subparagraph (C), including recommendations on the development of effective research methods for the study of drug safety questions.

"(C) ESTABLISHMENT OF THE POSTMARKET RISK IDENTIFICATION AND ANALYSIS SYSTEM.—

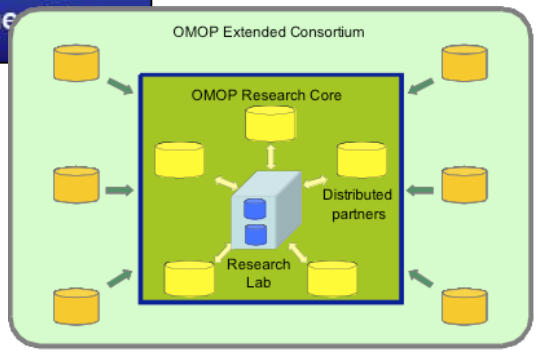


Risk Identification and Analysis System:

a systematic and reproducible process to efficiently generate evidence to support the characterization of the potential effects of medical products from across a network of disparate observational healthcare data sources

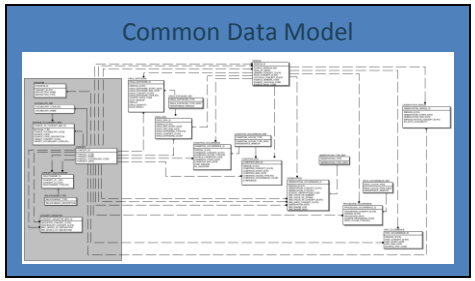
OMOP Experiment 1 (2009-2010)

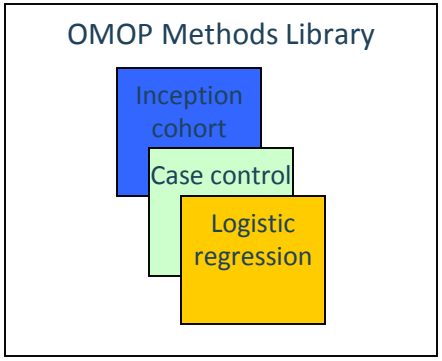
Observational Medical Outcomes Partne



- 10 data sources
- Claims and EHRs
- 200M+ lives

- Open-source
- Standards-based





- 14 methods
- Epidemiology designs
- Statistical approaches adapted for longitudinal data

Drug

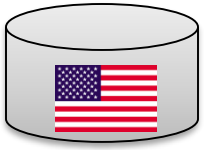
Outcome	ACE Inhibitors	Amphotericin B	Antibiotics: erythromycins, sulfonamides, tetracyclines	Antiepileptics: carbamazepine, phenytoin	Benzodiazepines	Beta blockers	Bisphosphonates: alendronate	Tricyclic antidepressants	Typical antipsychotics	Warfarin
Angioedema										
Aplastic Anemia										
Acute Liver Injury										
Bleeding										
Hip Fracture										
Hospitalization										
Myocardial Infarction										
Mortality after MI										
Renal Failure										
GI Ulcer Hospitalization										

Observational Medical Outcomes Partnership

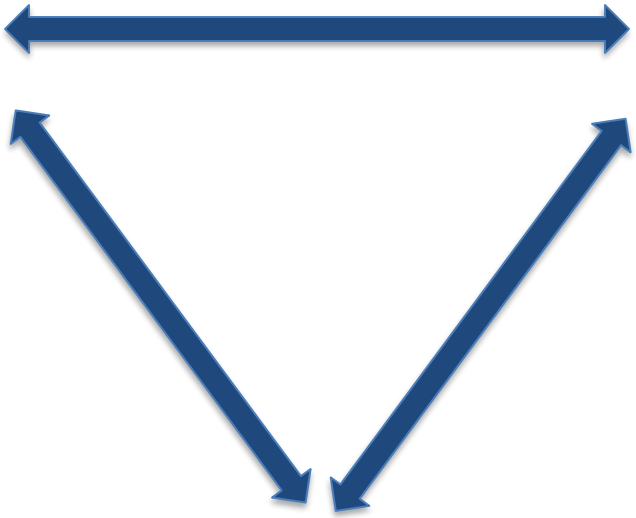
OMOP Experiment 2 (2011-2012)

Observational data

4 claims databases



1 ambulatory EMR



Methods

- Case-Control
- New User Cohort
- Disproportionality methods
- ICTPD
- LGPS
- Self-Controlled Cohort
- SCCS

Drug-outcome pairs

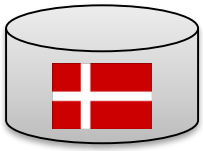
	Positives	Negatives
Total	165	234
Myocardial Infarction	36	66
Upper GI Bleed	24	67
Acute Liver Injury	81	37
Acute Renal Failure	24	64

Observational Medical Outcomes Partnership

European OMOP Experiment

Observational data

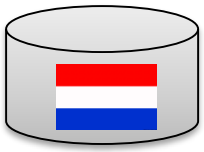
Aarhus



Pedianet

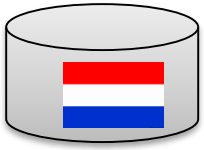
ARS

IPCI

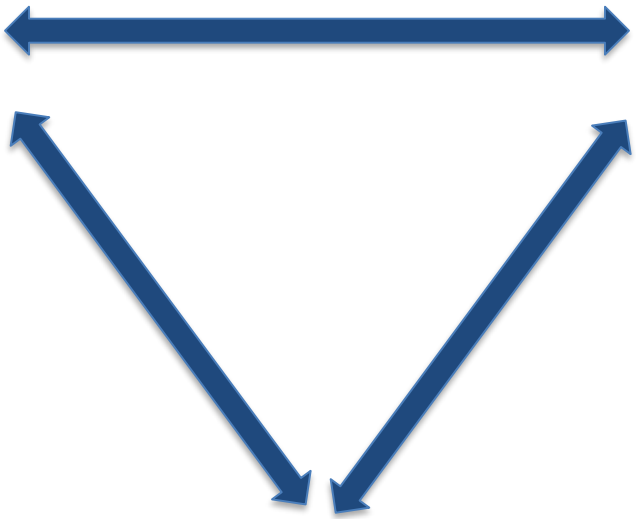


HS

PHARMO



eu-adr



Methods

- Case-Control
- New User Cohort
- Disproportionality methods
- ICTPD
- LGPS
- Self-Controlled Cohort
- SCCS

Drug-outcome pairs

	Positives	Negatives
Total	165	234
Myocardial Infarction	36	66
Upper GI Bleed	24	67
Acute Liver Injury	81	37
Acute Renal Failure	24	64

Observational Medical Outcomes Partnership

Ground Truth for OMOP Experiment

isoniazid

fluticasone

indomethacin

clindamycin

	Positive controls	Negative controls	Total
Acute Liver Injury	81	37	118
Acute Myocardial Infarction	36	66	102
Acute Renal Failure	24	64	88
Upper Gastrointestinal Bleeding	24	67	91
Total	165	234	399

ibuprofen

loratadine

sertraline

pioglitazone

Criteria for positive controls:

- Event listed in Boxed Warning or Warnings/Precautions section of active FDA structured product label
- Drug listed as 'causative agent' in Tisdale et al, 2010: Drug-Induced Diseases
- Literature review identified no powered studies with refuting evidence of effect

Criteria for negative controls:

- Event not listed anywhere in any section of active FDA structured product label
- Drug not listed as 'causative agent' in Tisdale et al, 2010: Drug-Induced Diseases
- Literature review identified no powered studies with evidence of potential positive association

Data Used in European Experiment

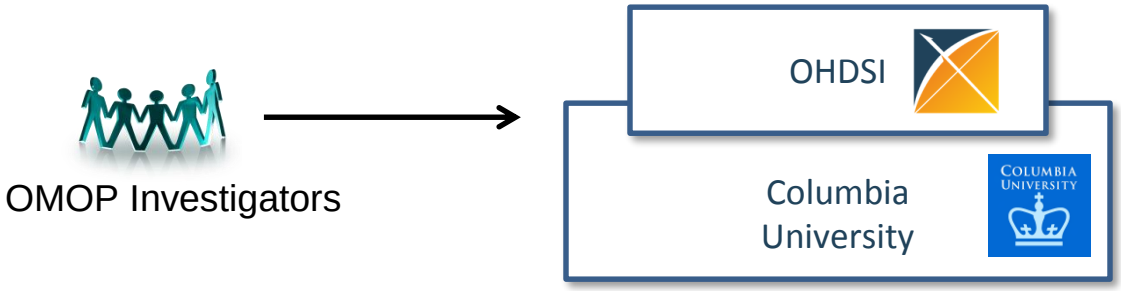
Name	Description	Population
Aarhus	Danish national health registry, covering the Aarhus region. Includes inhabitant registry, drug dispensations, hospital claims, lab values, and death registry.	2 M 
ARS	Italian record linkage system covering the Tuscany region, including inhabitant registry, drug dispensations, hospital claims, and death registry	4 M 
Health-Search	Italian general practice database (no children)	1 M 
IPCI	Dutch general practice database	0.75 M 
Pedianet	Italian general practice pediatric database	0.14 M 
PHARMO	Dutch record linkage system. Includes inhabitant registry, drug dispensations, hospital claims, and lab values.	1.28 M 

Observational Medical Outcomes Partnership

Main findings in OMOP experiment

- Heterogeneity in estimates due to choice of database
- Heterogeneity in estimates due to analysis choices
- Except little heterogeneity due to outcome definitions
- Good performance (AUC > 0.7) in distinguishing positive from negative controls for optimal methods when stratifying by outcome and restricting to powered test cases
- Self controlled methods perform best for all outcomes

OMOP – OHDSI的前世今生



- 健康观测数据科学和信息联盟（OHDSI）是一个多利益相关方，跨学科协作机构，旨在创建开源解决方案，通过大规模分析提供观测性健康数据的价值

- OHDSI由位于哥伦比亚大学的协调中心牵头，创建了一个由科研人员和健康观测数据库组成的国际化网络
  - 公开，开放
  - 无药厂资助
  - 国际化

http://ohdsi.org

OHDSI的展望

OHDSI合作者可访问拥有10亿患者的网络，以生成有关医疗保健各方面的证据。 世界各地的患者和临床医生以及其他决策者每天都使用OHDSI工具和证据。

加入我们的旅程

http://ohdsi.org

OHDSI: 一个全球化组织

OHDSI 合作者:

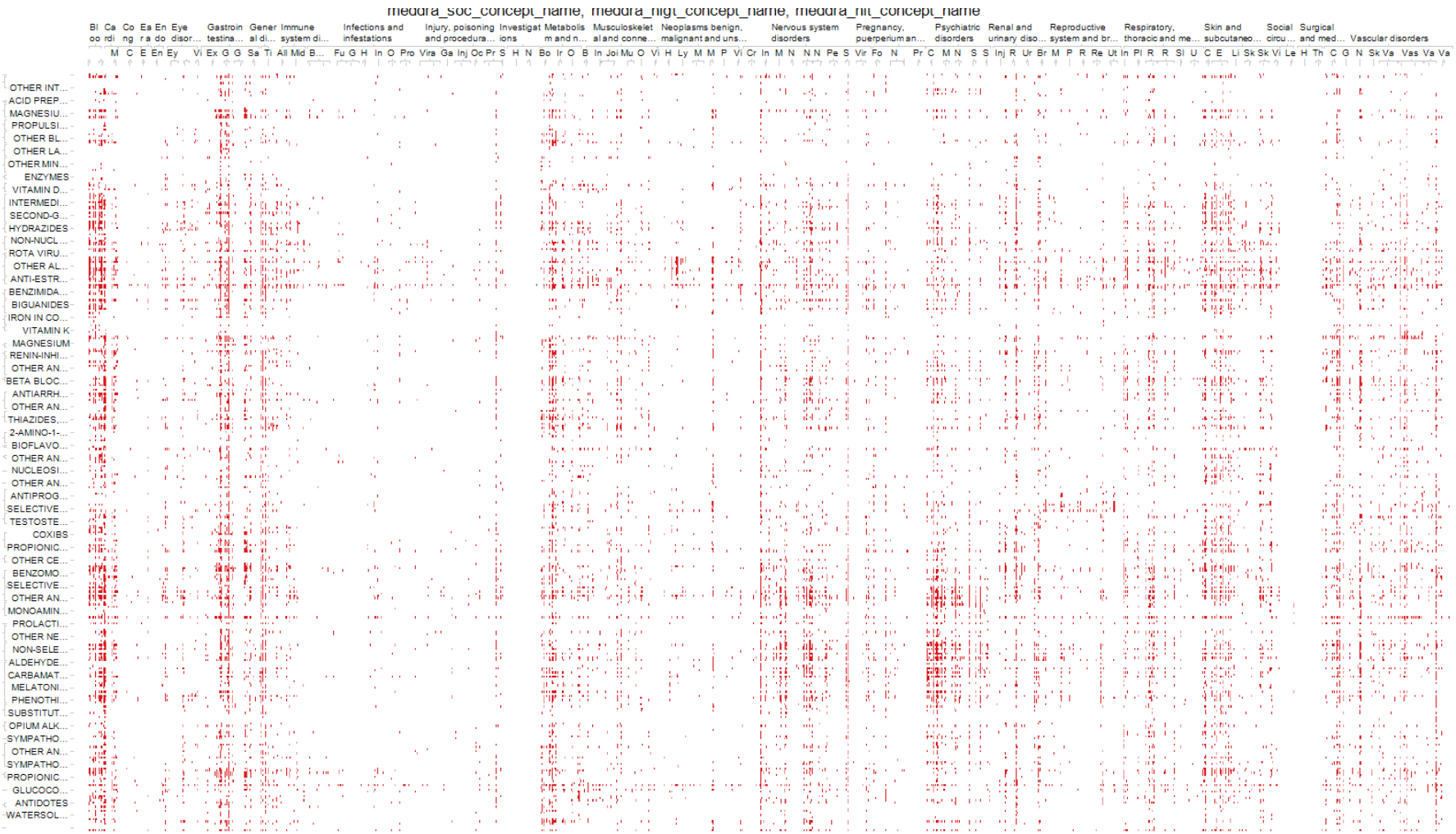
- >200 位来自学术界，政府和工业界的研究者
- >17 个国家

OHDSI 数据网络:

- >82 个数据库，来自于17个国家
- 12亿条病人记录(有重复)
- ~1.15 亿美国以外的患者

目前在医疗保健领域证据产生的速度

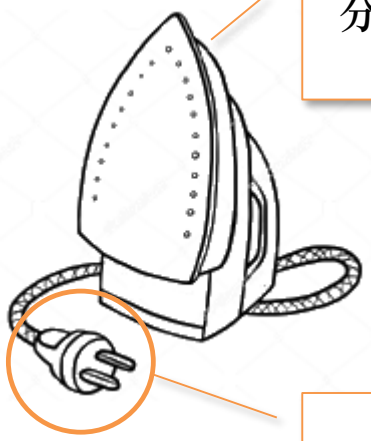
全部关注的结果



全部药物

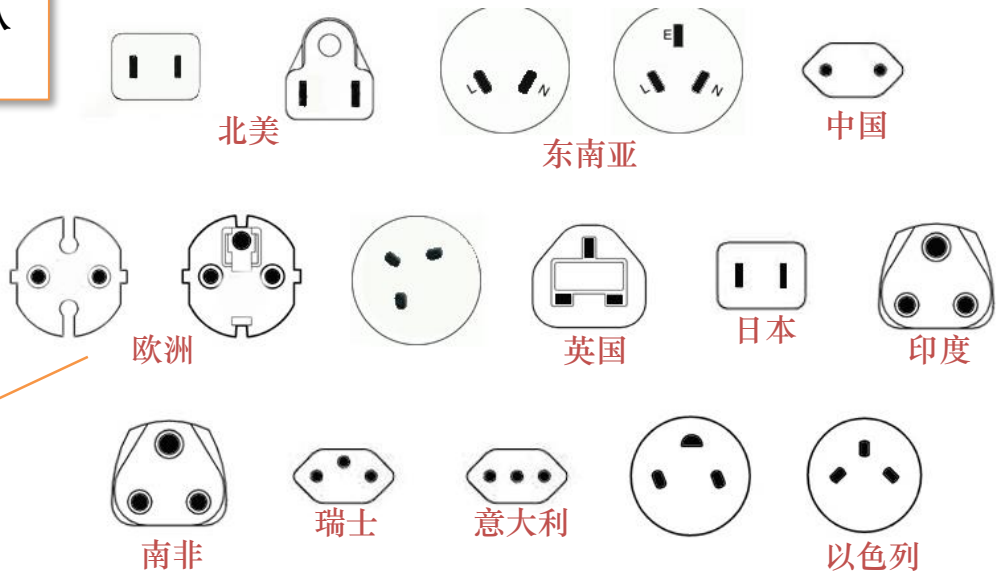
目前的方法: “一个研究 – 一个设计”

“在我的数据中对我的药物的依从性是怎样的?”

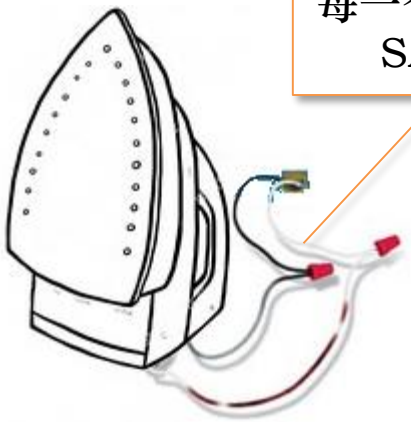


分析方法: 药物依从性

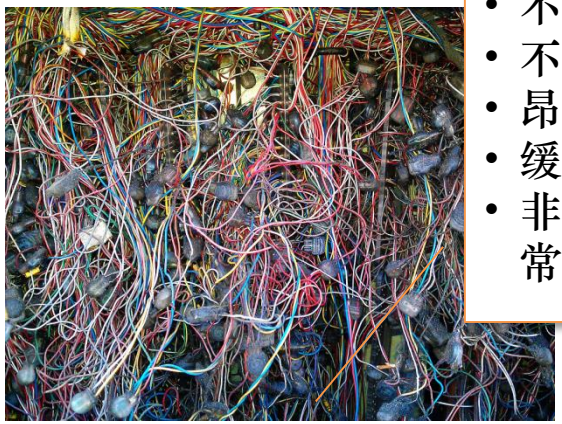
应用于数据



目前的方法:

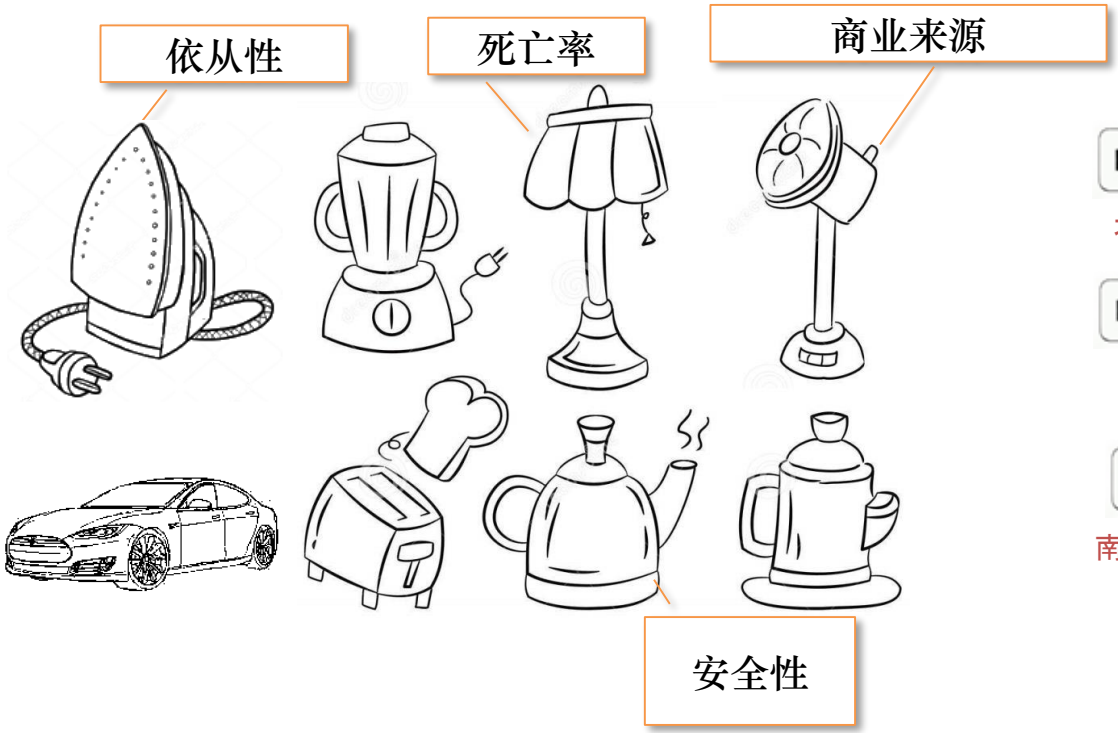


每一个研究都有一个SAS 或R脚本



- 不可扩展
- 不透明
- 昂贵
- 缓慢
- 非专业人员无法常规使用

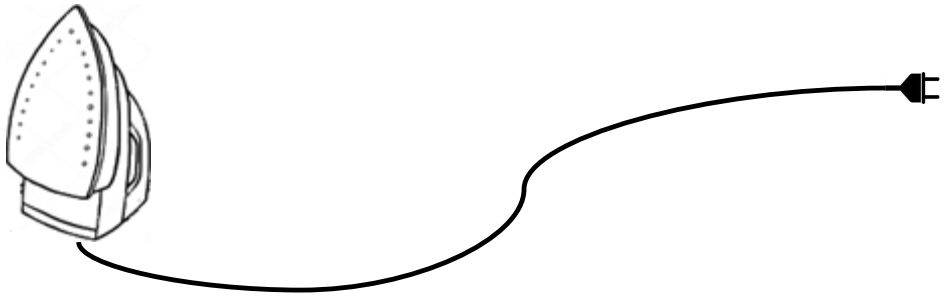
解决方案: 数据标准化以便于系统性研究



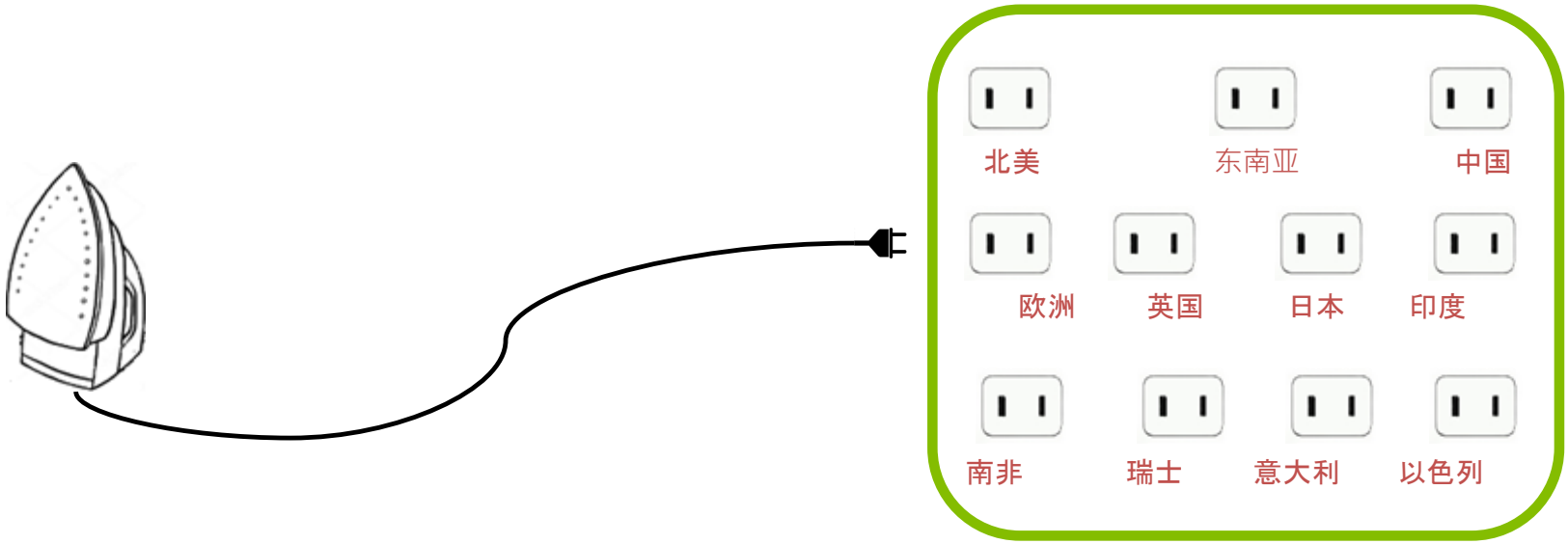
OHDSI 工具

OMOP CDM

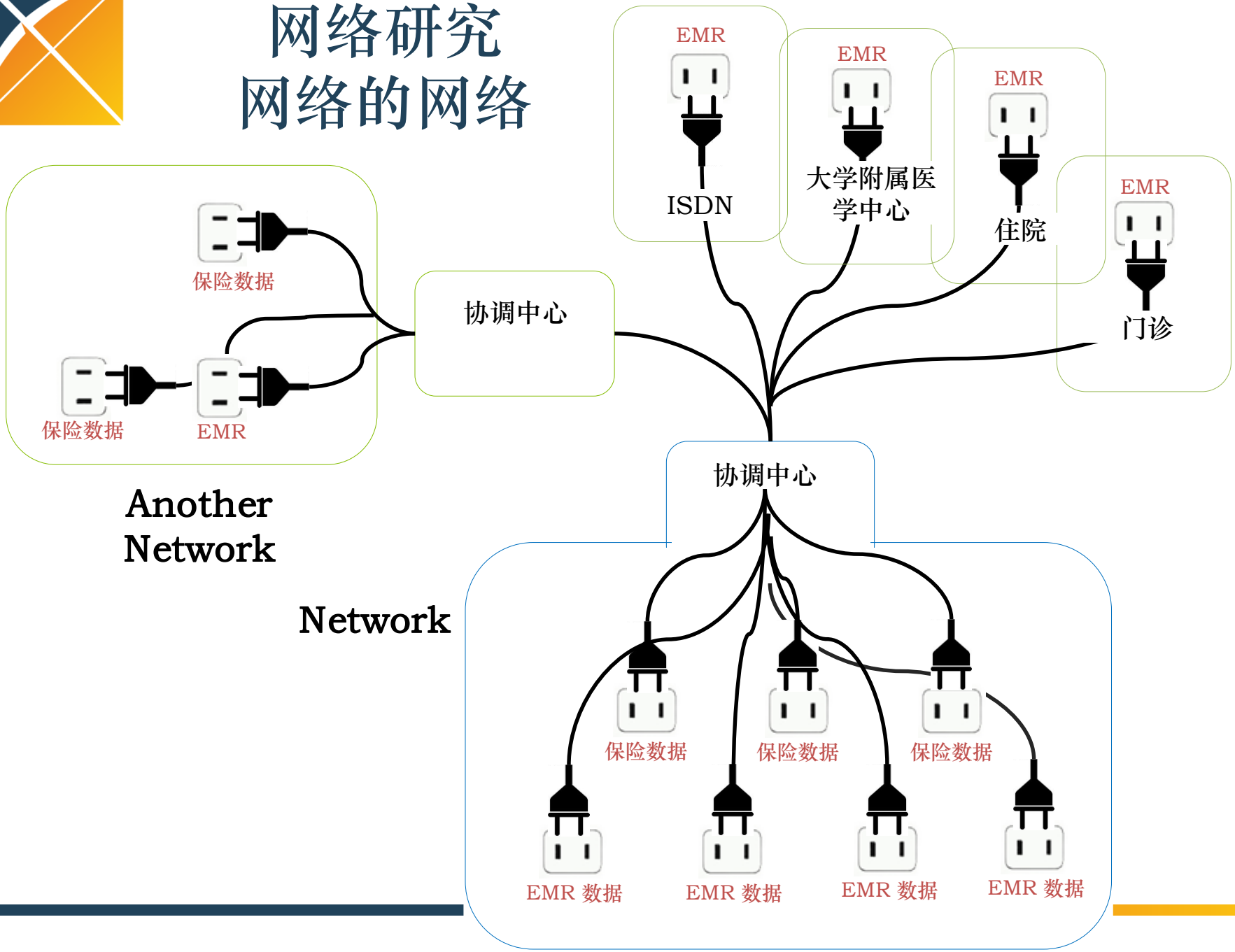
Analytics can be remote



Analytics can be behind hospital / data source firewalls



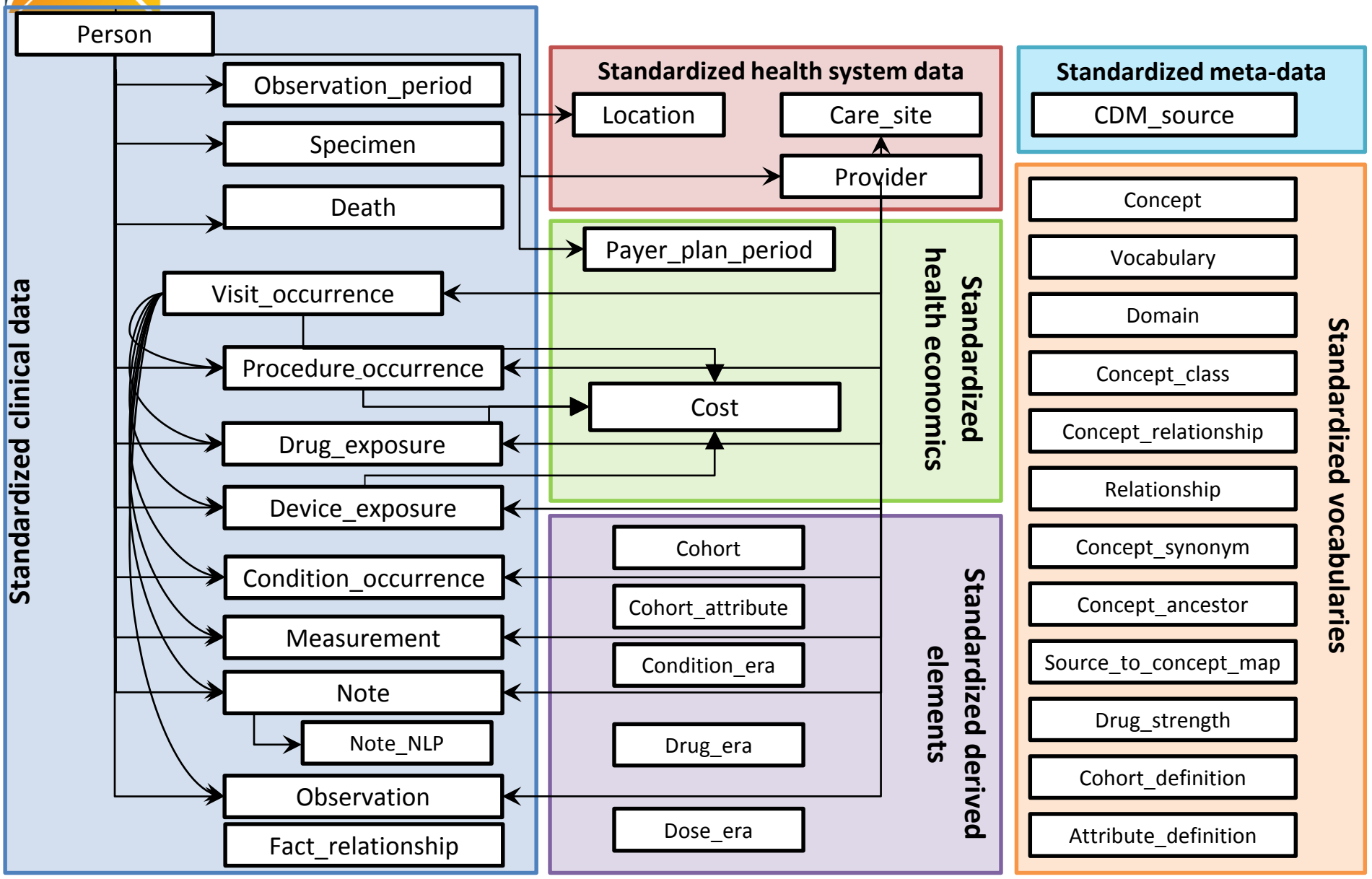
网络研究
网络的网络


EMR
ISDN
EMR
大学附属医学中心
EMR
住院
EMR
门诊
保险数据
保险数据
EMR
协调中心
Another Network
协调中心
Network
保险数据
保险数据
保险数据
EMR 数据
EMR 数据
EMR 数据
EMR 数据

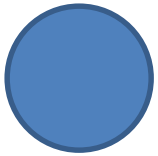
术语集

基本关系，祖先 & 后代在药品上的应用 SQL 范例

CDM 第5版核心版块

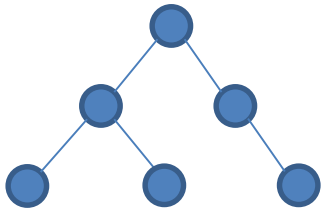

Standardized clinical data
Person
Observation_period
Specimen
Death
Visit_occurrence
Procedure_occurrence
Drug_exposure
Device_exposure
Condition_occurrence
Measurement
Note
Note_NLP
Observation
Fact_relationship
Standardized health system data
Location
Care_site
Provider
Standardized health economics
Payer_plan_period
Cost
Standardized derived elements
Cohort
Cohort_attribute
Condition_era
Drug_era
Dose_era
Standardized meta-data
CDM_source
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_relationship
Relationship
Concept_synonym
Concept_ancestor
Source_to_concept_map
Drug_strength
Cohort_definition
Attribute_definition

OMOP 术语集结构



全部内容: 所有概念均存于concept表

概念间的直接关系存于concept_relationship

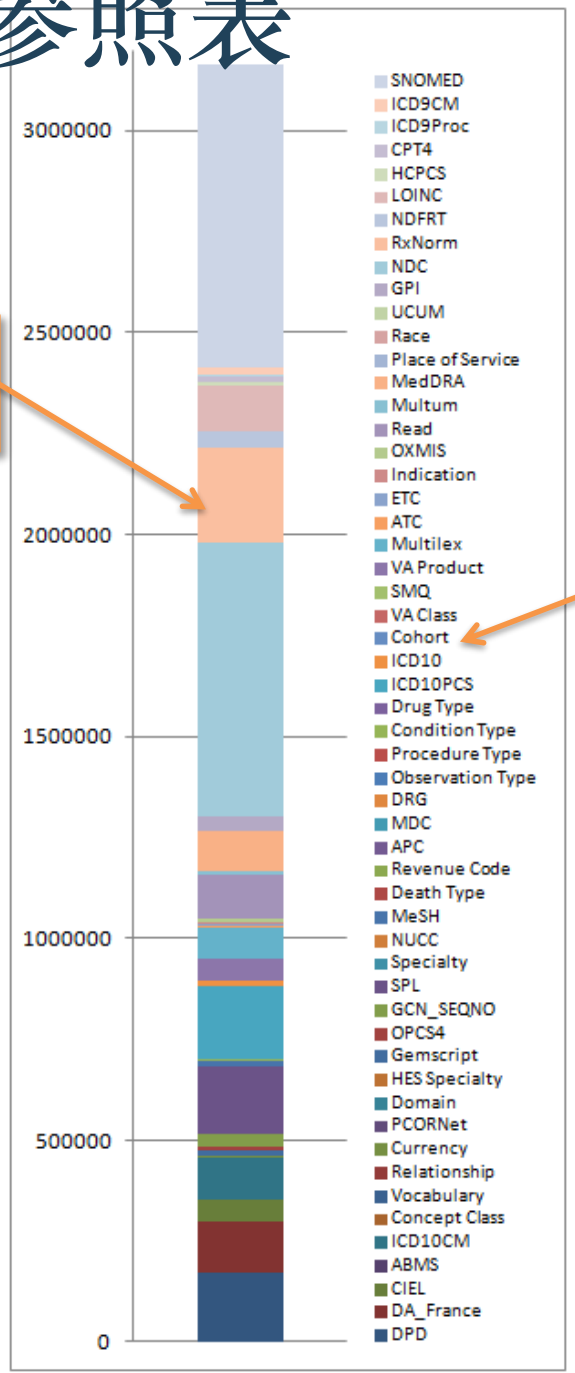


多级分层关系在concept_ancestor 预处理

单一概念参照表

所有术语汇总在一个表内

术语 ID



一条概念（Concept）储存的信息

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

CDM的全部内容都被编码为概念(Concepts)

- 每条概念均有唯一concept_id
- 所有细节均保存在**CONCEPT** 表:

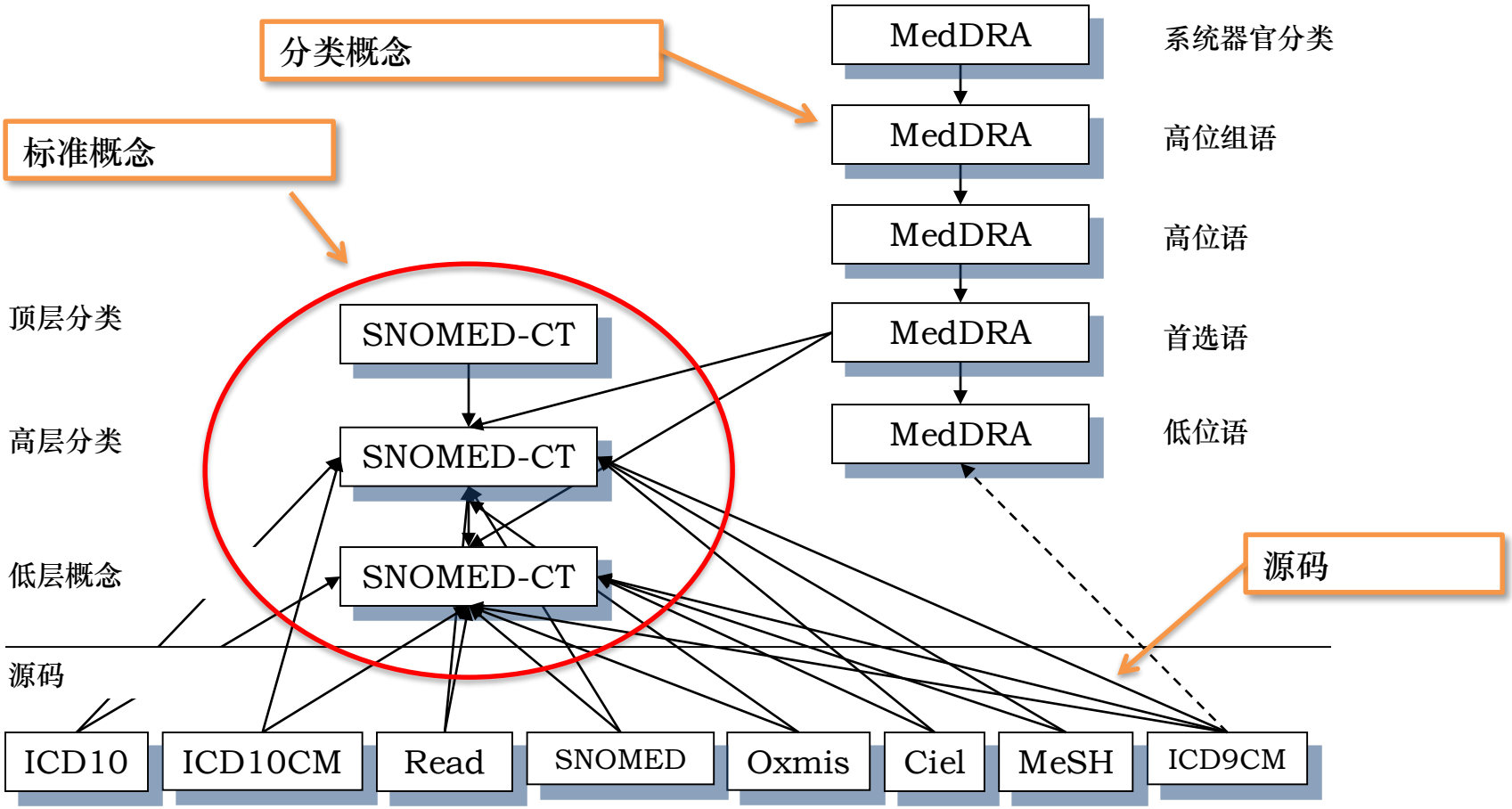
```
SELECT *

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	1/1/1970 0:00	12/31/2099 0:00	NULL

疾病概念


分类概念
标准概念
MedDRA
MedDRA
MedDRA
MedDRA
MedDRA
系统器官分类
高位组语
高位语
首选语
低位语
顶层分类
高层分类
低层概念
源码
SNOMED-CT
SNOMED-CT
SNOMED-CT
源码
ICD10
ICD10CM
Read
SNOMED
Oxmis
Ciel
MeSH
ICD9CM

查找正确的概念: #1

1. ..如果ID已知

```
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	

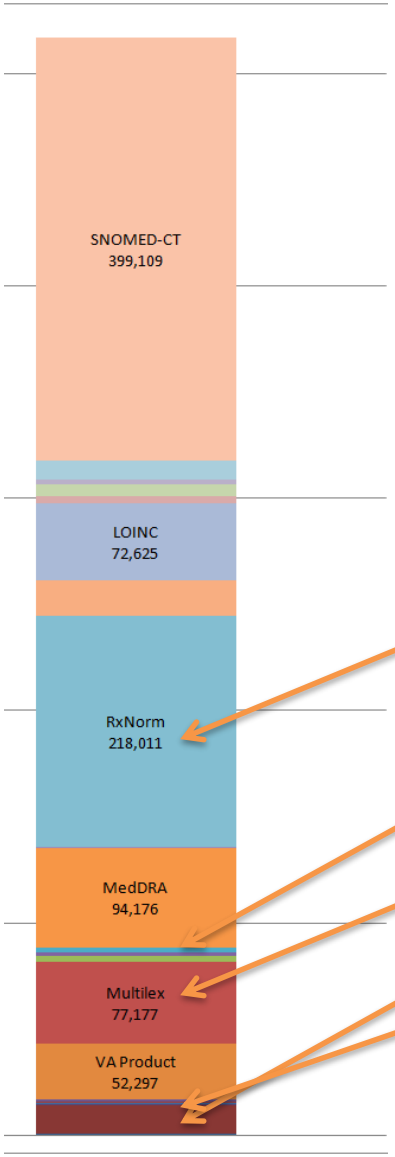
SNOMED code

2. ..如果编码（code）已知

```
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	

概念 ID versus 概念编码




```
SELECT *
FROM concept
WHERE concept_code = '1001';
```

同一编码

Concept_Name	Concept Class	Vocabulary_ID	Concept_Code
Antipyrine	Ingredient	RxNorm	1001
Aceprometazine maleate	Ingredient	BDPM	1001
Serum	Specimen	CIEL	1001
methixene hydrochloride	Ingredient	Multilex	1001
Brompheniramine Maleate, 10 mg/mL injectable solution	Multum	Multum	1001
ABBOTT COLD SORE BALM 4%/0.06% W/	Drug Product	LPD_Australia	1001
Residential Treatment - Psychiatric	Revenue Code	Revenue Code	1001

查找正确的概念: #2

3. ..如果名称（name）已知

```
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

查找正确的概念: #3

1. 只知道在另外一个术语集中对应的编码

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

ICD-9 不是标准概念

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;
```

不同术语集之间的映射

关系类型

CONCEPT _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	

映射 = 翻译

Step 1. 查找源概念（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. 标准化

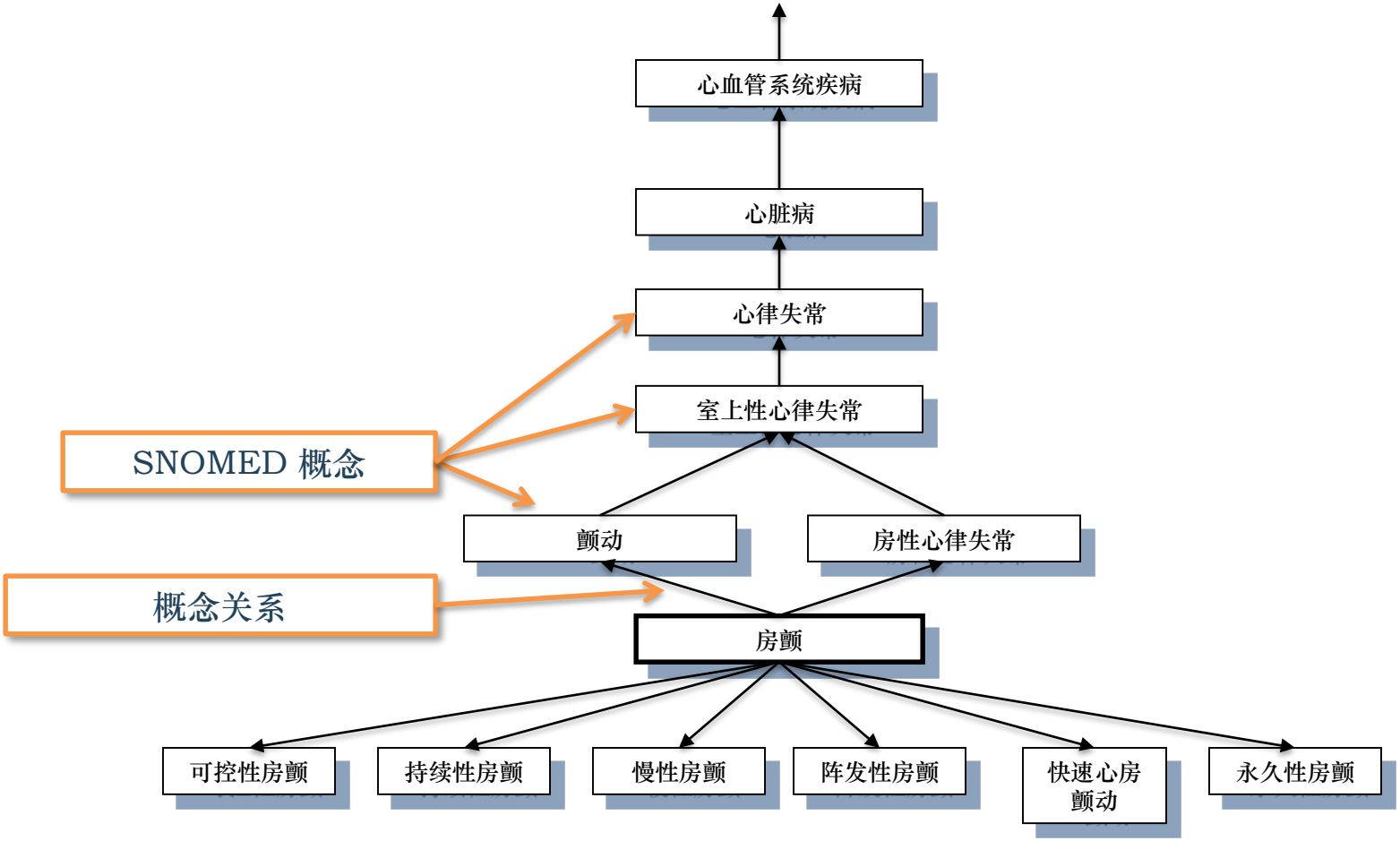
```
SELECT * FROM concept_relationship WHERE concept_id_1 = 44821957
    AND relationship_id = 'Maps to';
```

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. 查询翻译后的概念

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

理由 #2: 疾病层级


心血管系统疾病
心脏病
心律失常
室上性心律失常
SNOMED 概念
颤动
房性心律失常
概念关系
房颤
可控性房颤
持续性房颤
慢性房颤
阵发性房颤
快速心房颤动
永久性房颤

关系表

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

相关的概念

关系 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

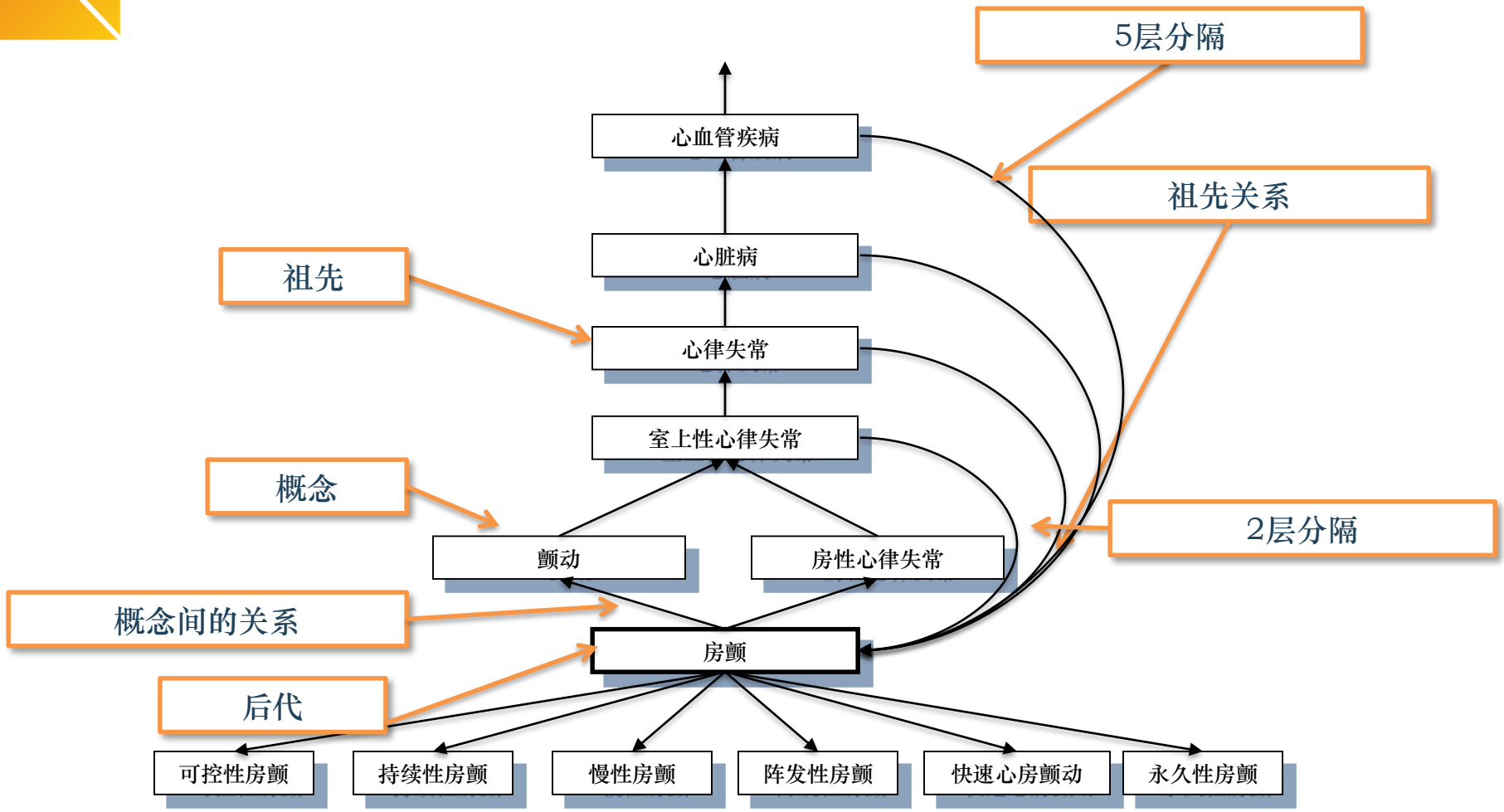
关系表 #2

```
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;
```

查找相关概念

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	H/O: atrial fibrillation	Observation	SNOMED	Context-dependent	S	312442005	1/1/1970 0:00	12/31/2099 0:00	NULL
Asso finding of	4203375	Family history of atrial fibrillation	Observation	SNOMED	Context-dependent	S	433276002	1/31/2009 0:00	12/31/2099 0:00	NULL
Asso finding of	42689685	Atrial fibrillation not detected	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 excluded	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	4139517	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	42709991	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

祖先关系(Ancestors): 更高层次的关系



一个概念的祖先

```
SELECT max_levels_of_separation, c.*
FROM concept_ancestor ca, concept c
WHERE ca.descendant_concept_id = 313217 /* Atrial fibrillation */
  AND ca.ancestor_concept_id = c.concept_id
ORDER BY max_levels_of_separation
;
```

作为后代

max_levels_of_separation	concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept
0	313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S
0	35204953	Atrial fibrillation	Condition	MedDRA	PT	C
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

一个概念的后代

作为祖先

```
SELECT max_levels_of_separation, c.*
FROM concept_ancestor ca, concept c
WHERE ca.ancestor_concept_id = 44784217 /* cardiac arrythmia */
  AND ca.descendant_concept_id = c.concept_id
ORDER BY max_levels_of_separation
;
```

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

查找上消化道出血 （Upper Gastrointestinal Bleeding)

1. 找到一些初始的概念

```
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. 查找标准概念

```
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

沿层级向上查找: 找到正确的概念

作为后代

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

向下查找: 检查内容是否正确

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

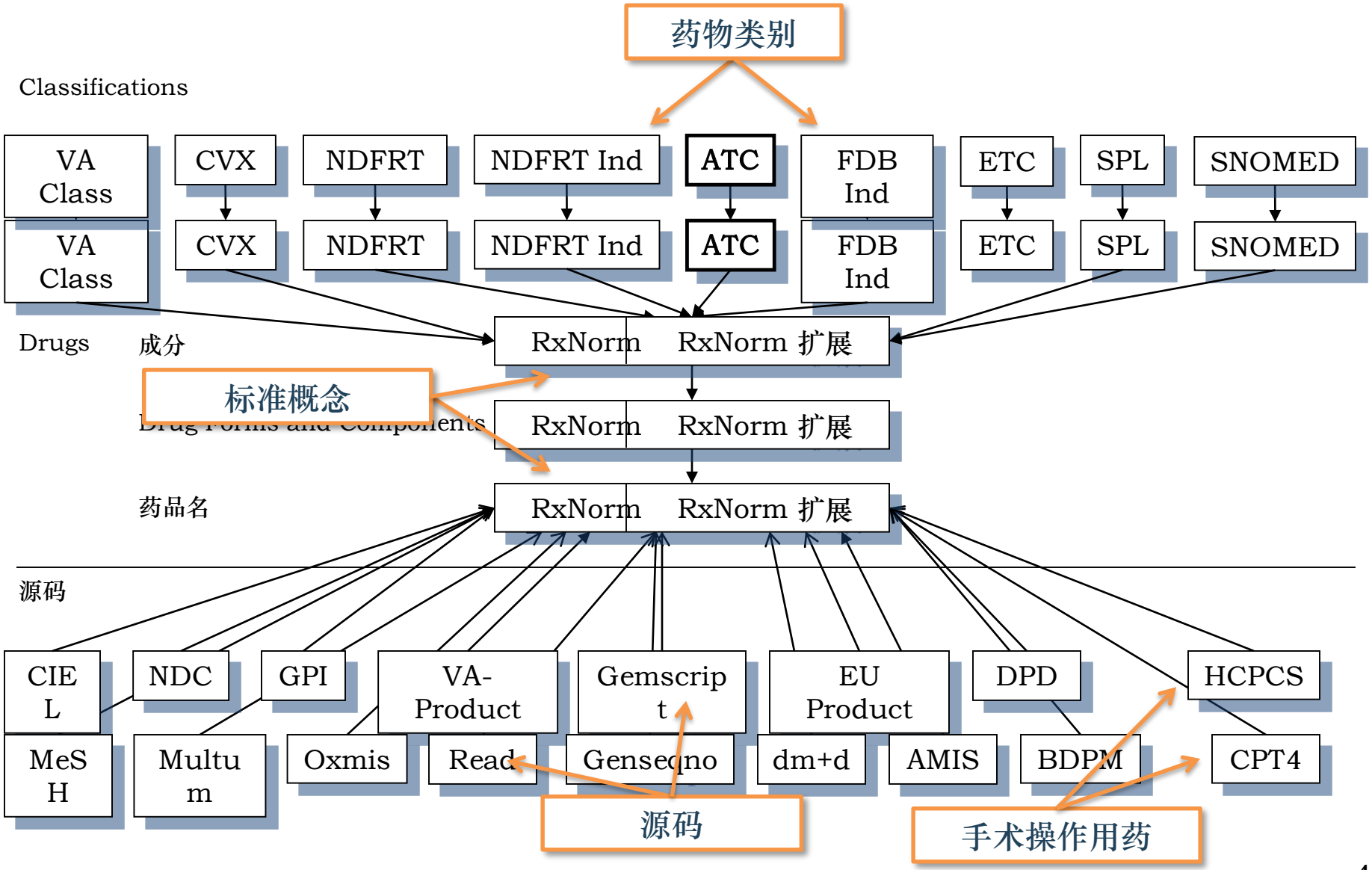
max_levels_of_separation	concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code
0	4291649	Upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	37372002
1	4318535	Duodenal hemorrhage					
1	23245	Esophageal bleeding					
1	4308202	Acute upper gastrointestinal 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

概念 4291649 及其所有的后代概念均为上消化道出血

对于药品一样适用吗?

- 编码
  - NDC, GPI, Multilex, HCPCS, etc.
- 概念
  - 药品名(通用名 及 商品名)
  - 药物成分
  - 药物类别
- 关系
- 祖先

药品层级


药物类别
Classifications
VA Class
CVX
NDFRT
NDFRT Ind
ATC
FDB Ind
ETC
SPL
SNOMED
VA Class
CVX
NDFRT
NDFRT Ind
ATC
FDB Ind
ETC
SPL
SNOMED
Drugs
成分
RxNorm
RxNorm 扩展
标准概念
RxNorm
RxNorm 扩展
药品名
RxNorm
RxNorm 扩展
源码
CIEL
NDC
GPI
VA-Product
Gemscript
EU Product
DPD
HCPCS
MeSH
Multum
Oxmis
Read
Gensequo
dm+d
AMIS
BDPM
CPT4
源码
手术操作用药

查找华法林（warfarin)

1．通过关键字查找活性化合物华法林

```
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

查找氯吡格雷（Clopidogrel）

1. 通过NDC代码查找含有氯吡格雷的药品名:

   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

查找氯吡格雷的成分

2. 查找成分是氯吡格雷的药品的祖先

```
SELECT a.max_levels_of_separation, c.*
FROM concept_ancestor ca
JOIN concept c ON ca.ancestor_concept_id = c.concept_id
WHERE ca.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

氯吡格雷

药品类别

查找成分

3. 查询后代(其他包含华法林和氯吡格雷的药品)

```
SELECT max_levels_of_separation, c.*
FROM concept_ancestor ca
JOIN concept c ON ca.descendant_concept_id = c.concept_id
WHERE ca.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

查找某一药品类别下的成员

4. 查找抗凝血药类别下属于药品成分（ingredient descendant)类别的后代

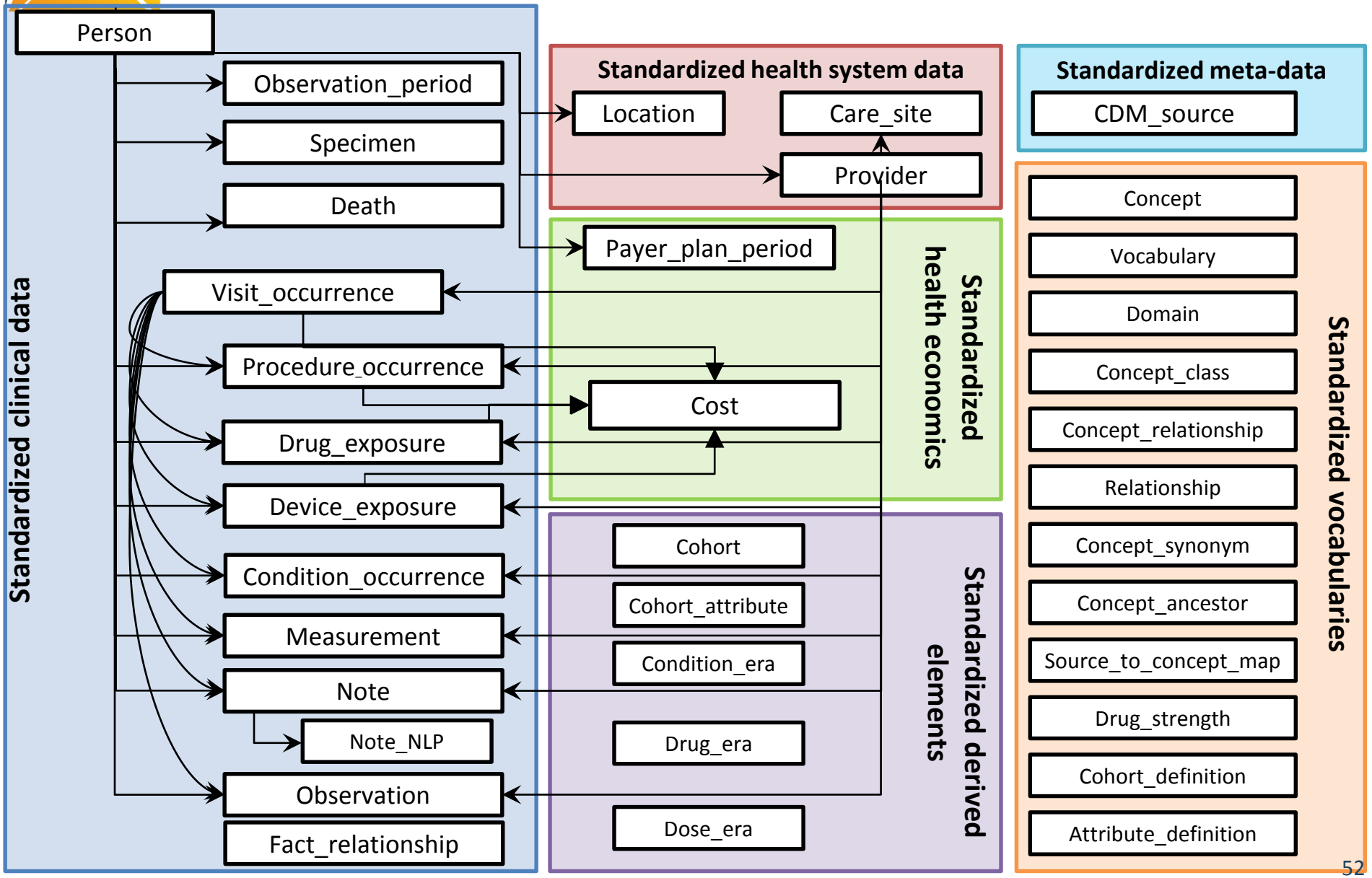
```
SELECT max_levels_of_separation，c.*
FROM concept_ancestor ca
JOIN concept c ON ca.descendant_concept_id = c.concept_id
WHERE ca.ancestor_concept_id = 21600961 /* ATC Antithromboic Agent */
  AND c.concept_class_id = 'Ingredient'
ORDER BY max_levels_of_separation;
```

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
19129274	reviparin	Drug	RxNorm	Ingredient

通用数据模型

模型的历史
深入讨论
Era 讨论

CDM 第5版核心版块


Standardized clinical data
Person
Observation_period
Specimen
Death
Visit_occurrence
Procedure_occurrence
Drug_exposure
Device_exposure
Condition_occurrence
Measurement
Note
Note_NLP
Observation
Fact_relationship
Standardized health system data
Location
Care_site
Provider
Standardized health economics
Payer_plan_period
Cost
Standardized derived elements
Cohort
Cohort_attribute
Condition_era
Drug_era
Dose_era
Standardized meta-data
CDM_source
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_relationship
Relationship
Concept_synonym
Concept_ancestor
Source_to_concept_map
Drug_strength
Cohort_definition
Attribute_definition

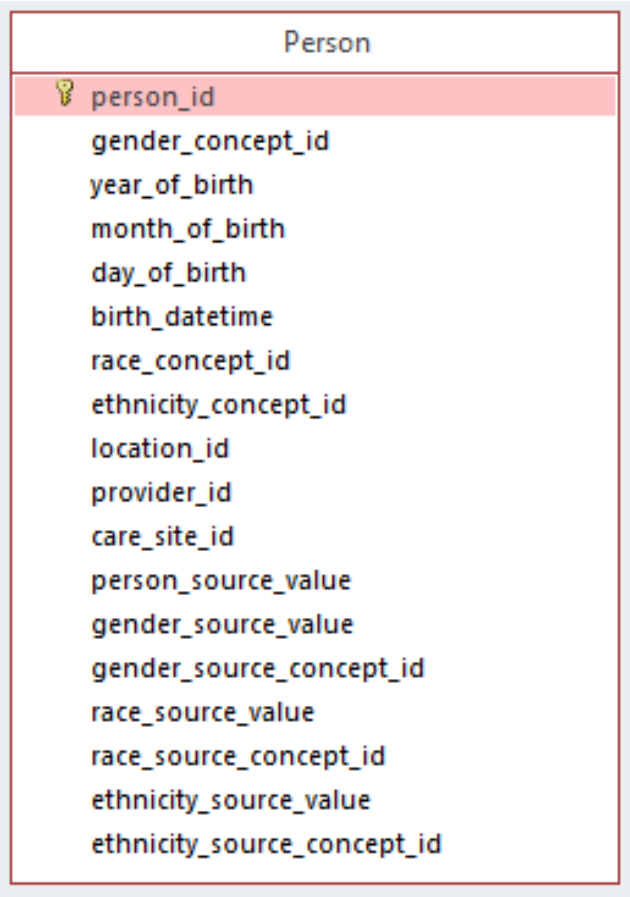
OMOP CDM Principles

- OMOP model is an information model
  - Vocabulary(Conceptual) and Data Model are blended
  - Domain-oriented concepts
- Patient centric
- Accommodates data from various sources
- Preserves data provenance
- Extendable
- Evolving

OMOP CDM Standard Domain Features

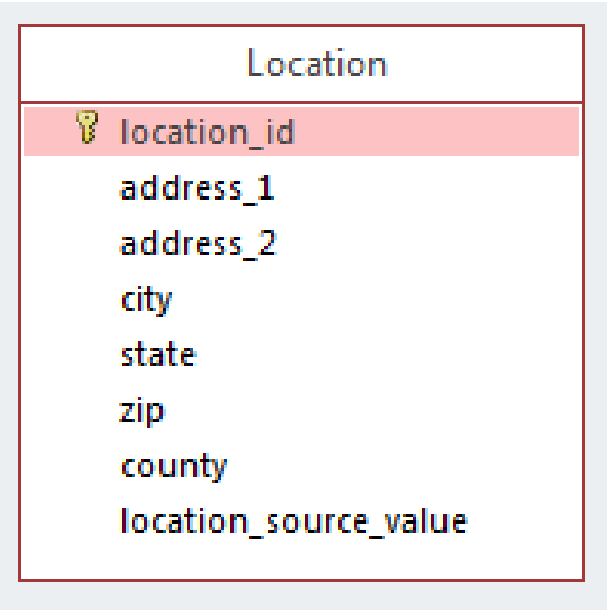
Feature	Description and purpose	Field name convention	Example
Patient centric	Every domain table has **patient identifier**. Patient data can be retrieved independently from other domains.	**person_id**	person_id 123
Unique domain identifier	Every domain table has a unique primary key to identify domain **entities**	<entity>**_id**	condition_occurrence_id 470985
Standard concept from a respective vocabulary domain	Integration with the vocabulary. Foreign key into the Standard Vocabulary for **Standard Concept**	<entity>**_concept_id**	condition_concept_id 313217 (SNOMED "Atrial Fibrillation")
Source concept from a respective vocabulary domain	Provenance. Foreign key into the Standard Vocabulary for **Source Concept**	<entity>**_source_concept_id**	condition_source_concept_id 44821957 (ICD9CM "Atrial Fibrillation")
Source value	Provenance. Verbatim information from the source data, **not to be used** by any standard analytics	<entity>**_source_value**	condition_source_value 427.31 (ICD9CM "Atrial Fibrillation")
Source type	Provenance. Foreign key into the Vocabulary for the **origin of the**	<entity>**_type_concept_id**	condition_type_concept_id 38000199 ("Inpatient header – primary")

PERSON表



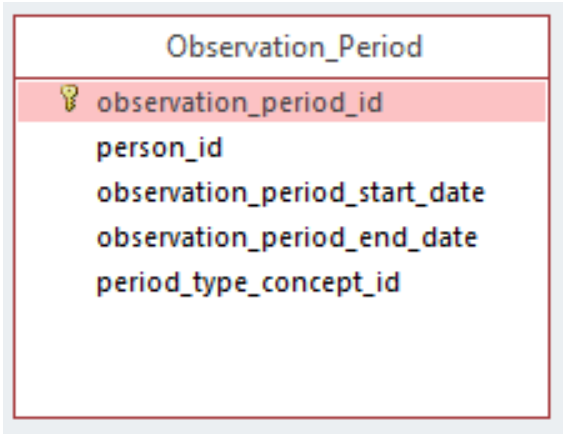
- 每个人要有唯一的记录
- 性别，种族使用术语集: HL7 标准
- 地理位置/人口统计学若无历史资料: 选择最近一次的数据
- 位置特点: LOCATION表的外键。Location表为每一个不同的地点保存一条记录
- 出生年费必填，日期/月份可选

LOCATION表



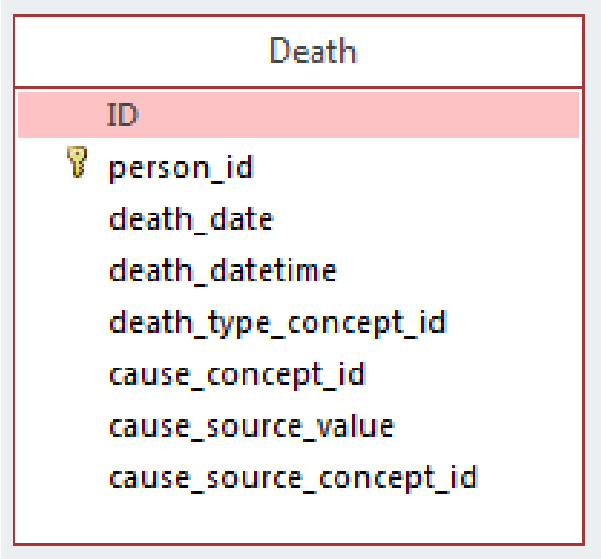
- 每个地理位置有一条记录
- 地理位置的记录随数据源的不同而高度不一致，目前用处有限

OBSERVATION_PERIOD表



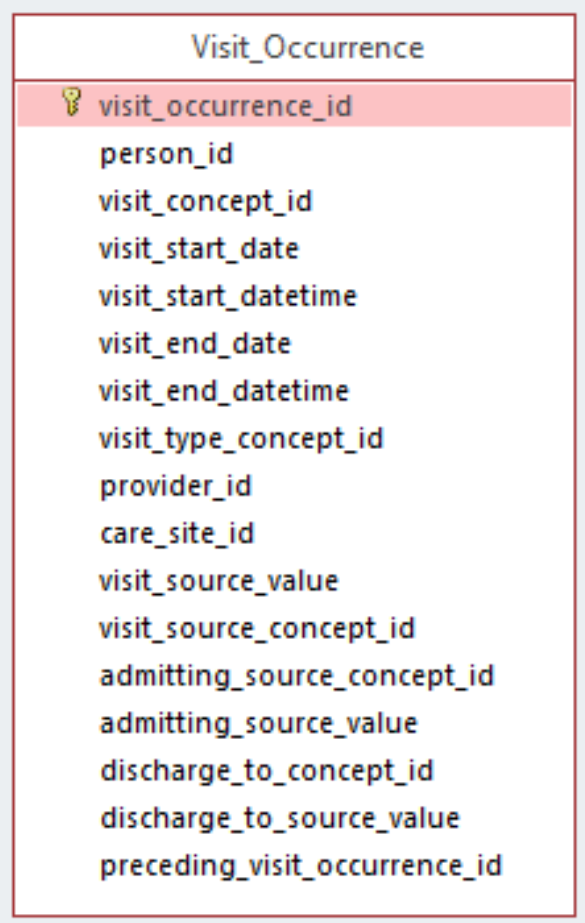
- 记录数据源收集数据的一段时间
- 分析方法必须的信息
- 如果在收集数据时有中断，一个人可以有多条（观察时间段）记录
- 挑战：观察时间段由源数据决定

DEATH表



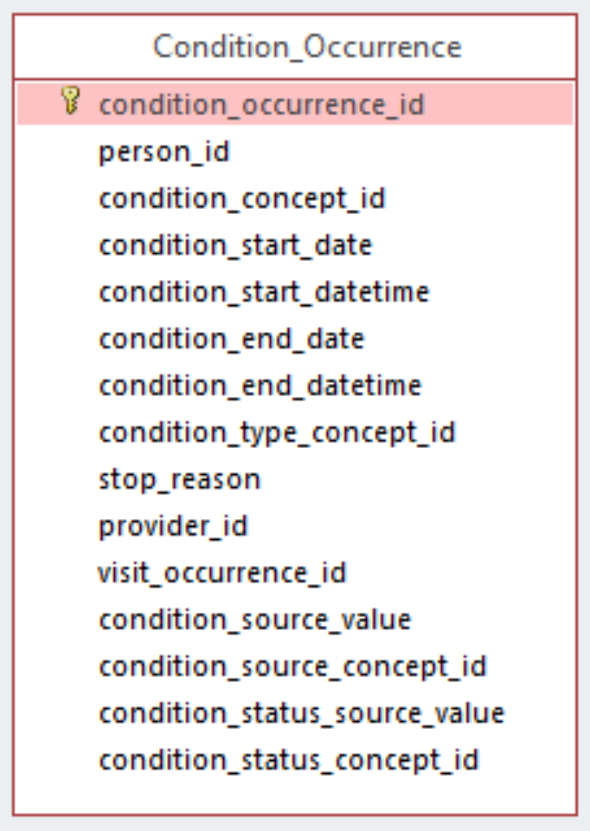
- 死亡记录可以无死因
- 一个人只允许有一条死亡记录

VISIT_OCCURRENCE表



- 就诊 visit <> 记录‘encounter’:
  - –多次就诊经常需要综合以避免重复计算保险
  - –不包括住院病人转诊
- 就诊类别
  - –住院
  - –急诊
  - –住院/急诊
  - –门诊
  - –长期护理
- 术语集: OMOP
- 其他属性: 就诊开始时间/结束时间, 医生, 入院来源，出院处置

CONDITION_OCCURRENCE表



- 术语集: SNOMED -> 分类
- **数据来源:**
  - 计费诊断 （billing diagnosis）(住院，门诊)
  - **疾病列表** （problem list）
- **一条记录<> 一次患病**

‘麻烦’ 的情况

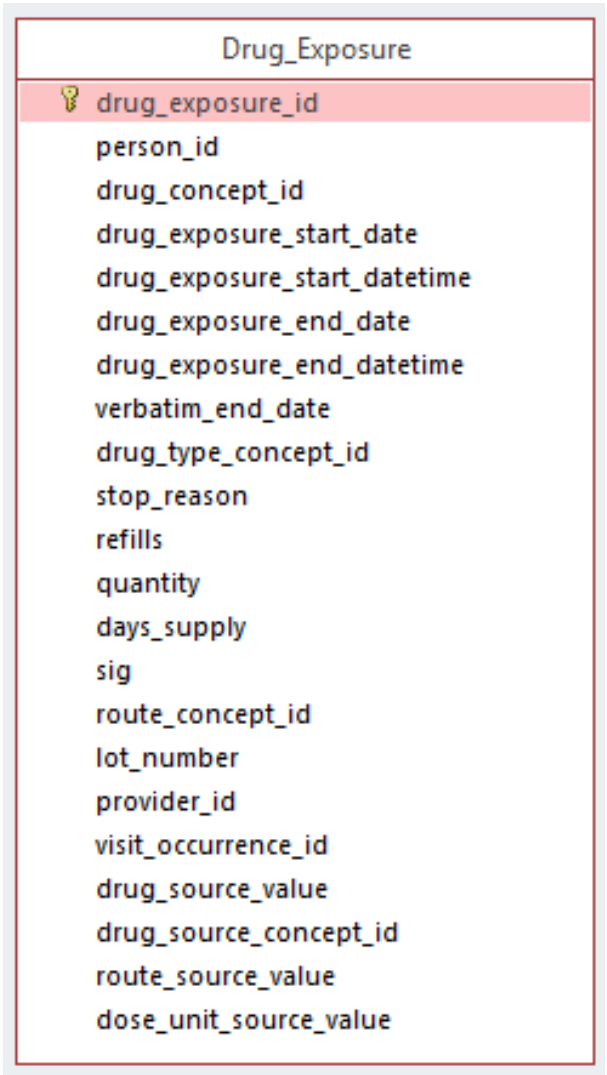
编码映射后到的版块与源版块不一致

Description	AUT MESUR CHIMIO PROPHYL
CIM10	0Z2920
CIM10 Description	OTH PROPHYLACTIC CHEMO
Maps to	DRUG

Description	EXSPEC DEPS AUT MAL PREC
CIM10	0Z1380
CIM10 Description	SPEC SCR OTHER SPEC DIS
Maps to	OBSERVATION

Description	GRSS CONSTT FORTUITMT
CIM10	0Z3300
CIM10 Description	PREG STATE INCIDENT
Maps to	CONDITION

DRUG_EXPOSURE表

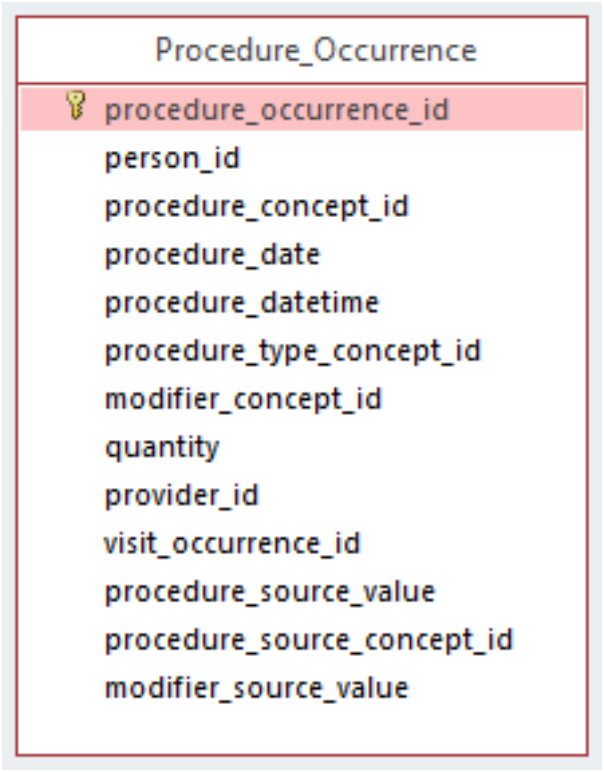


- 术语集: RxNorm-> 依据诊断和药品类别分类
- 数据源:
  - 药房配药
  - 处方
  - 用药史
- 来源的领域可能不同，所以药物暴露结束时间的推断也因此可能不同

‘麻烦的’ 药

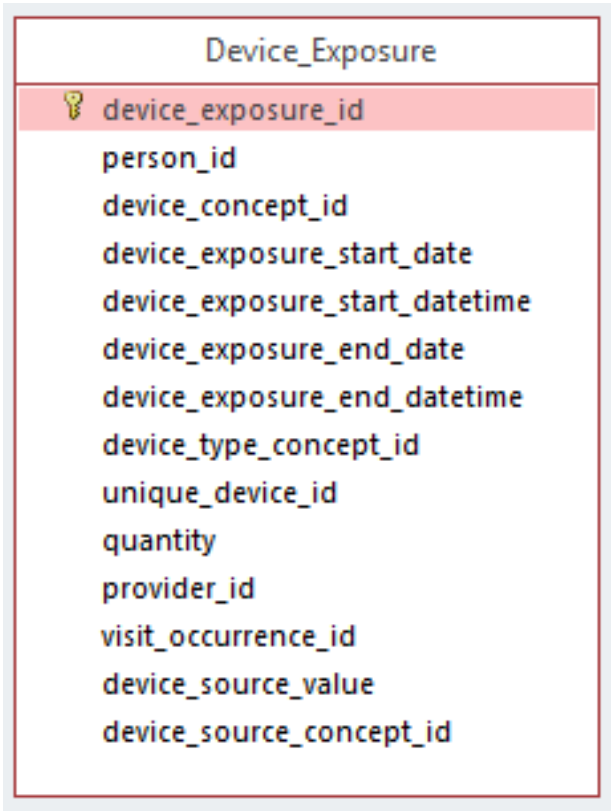
Drug Source Description	Form Desc	Admin Route Description	Generic Name	Maps To
OBSERVATION PERIOD	Miscellaneous	Unspecified	Documentation	OBSERVATION
LUMBAR DDS BELT	Miscellaneous	Unspecified	Unspecified	DEVICE
JOBST KNEE HIGH COMPRESSION STOCKING	Miscellaneous	Unspecified	Antiembolism stockings	DEVICE
MASKS	Miscellaneous	Unspecified	Masks	DEVICE
PEN NEEDLES	Miscellaneous	Unspecified	Needle	DEVICE

PROCEDURE_OCCURRENCE 表



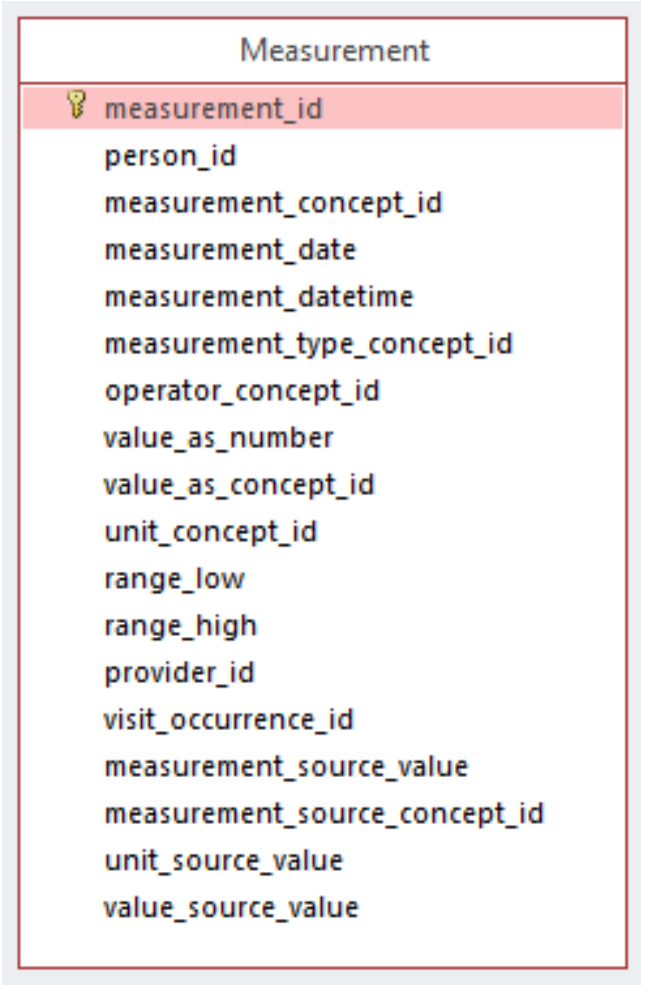
- 术语集: CPT-4, HCPCS, ICD-9 Procedures, ICD-10 Procedures, LOINC, SNOMED
- 手术操的术语集标准化程度最低，因此导致了一些重复记录

DEVICE_EXPOSURE表



- OMOP CDM 是唯一一个支持设备的数据模型
- OMOP CDM 可以容纳FDA 唯一的设备标识符（UDI），尽管大部分数据源甚至还没有纳入

MEASUREMENT表



- 实体-属性-值 （EAV）设计
- 术语集: LOINC, SNOMED
- 数据源: 结构化的，量化的测量指标，比如实验室检查
- 测量单位
  - 测量单位术语集: UCUM
- 测量结果不允许自由格式

测量（Measurement)数据的问题

- 源数据的测量单位不统一
  - –使评价和研究难以进行

```
select distinct unit_source_value
from measurement
where measurement_concept_id IN
 (
   SELECT concept_id
   FROM concept
   WHERE concept_name like '%LDL%'
     AND standard_concept = 'S'
     AND domain_id = 'Measurement'
 )
```

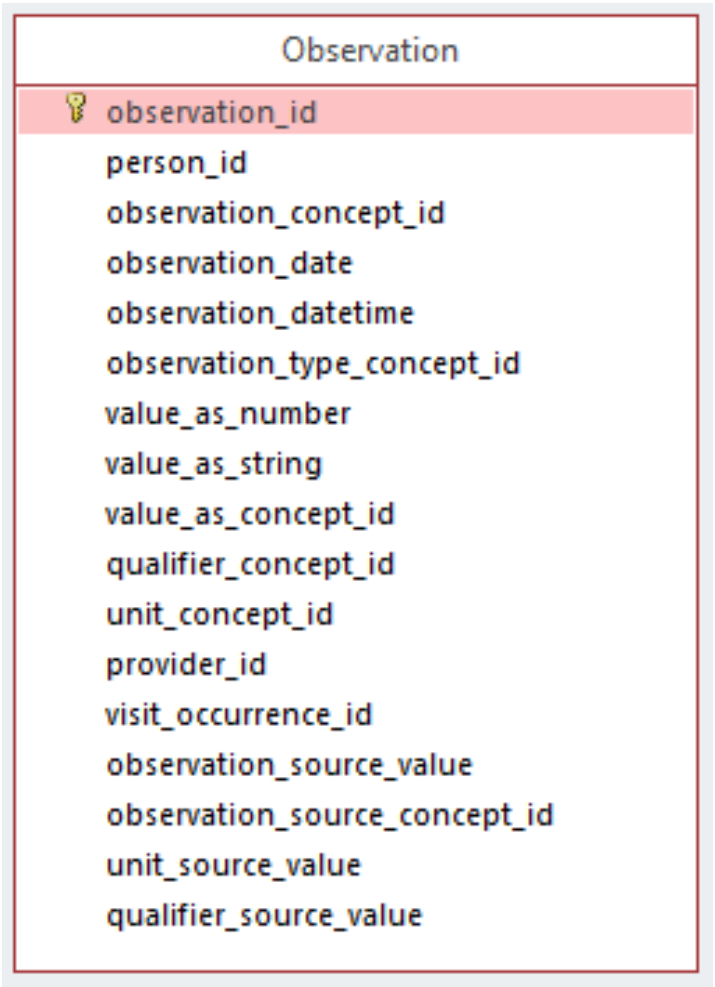
unit_source_value
N < 3.5
g/l
NULL
< 3,2
ng/ml
mmol/l
N < 3.6

测量（Measurement)数据的问题

```
select distinct(round (value_as_number/10))*10 as value,
       count (*)
from measurement
where measurement_concept_id in
 (
   SELECT concept_id
   FROM concept
   WHERE concept_name like '%LDL%'
      AND standard_concept = 'S'
      AND domain_id = 'Measurement'
  )
group by value
```

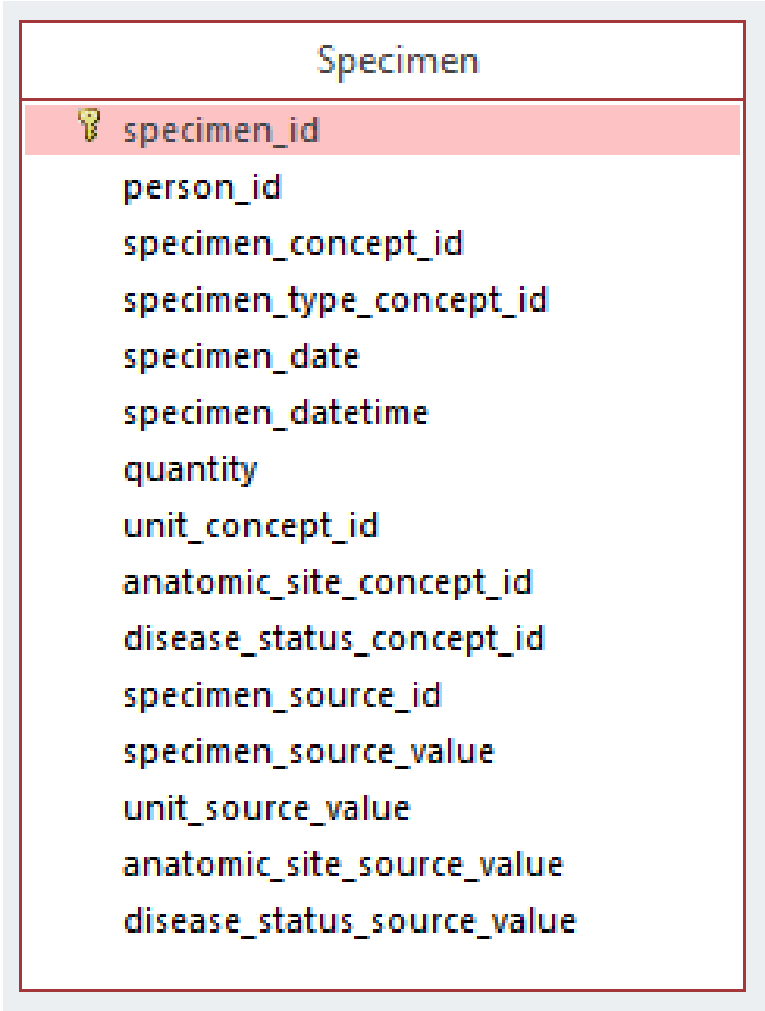
value	count
0	39784
10	302
180	1
20	1
120	1
30	2
NULL	2
4380	1
50	2
60	1
430	1

OBSERVATION表



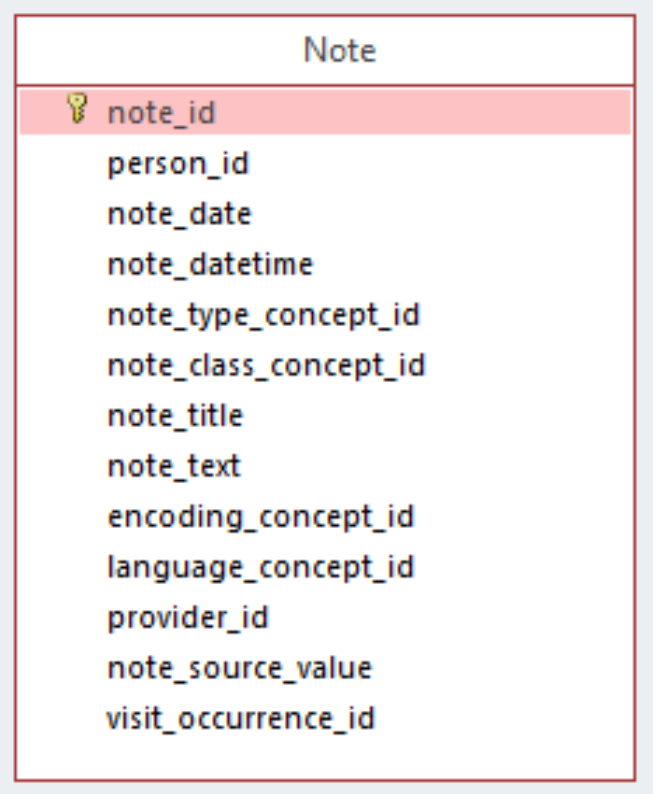
- “全覆盖”型EAV 设计以便于收集所有其他数据:
  - 观察: ‘问题’
  - 值: ‘回答’
    - 可以是数字，概念，或者字符串 (e.g. 纯文字)
- CDM扩展工具，“游乐场”
- 不是所有的“问题”都是标准化的，源数据可以生成“自定义”的观察（对于登记处数据尤其重要）

SPECIMEN表



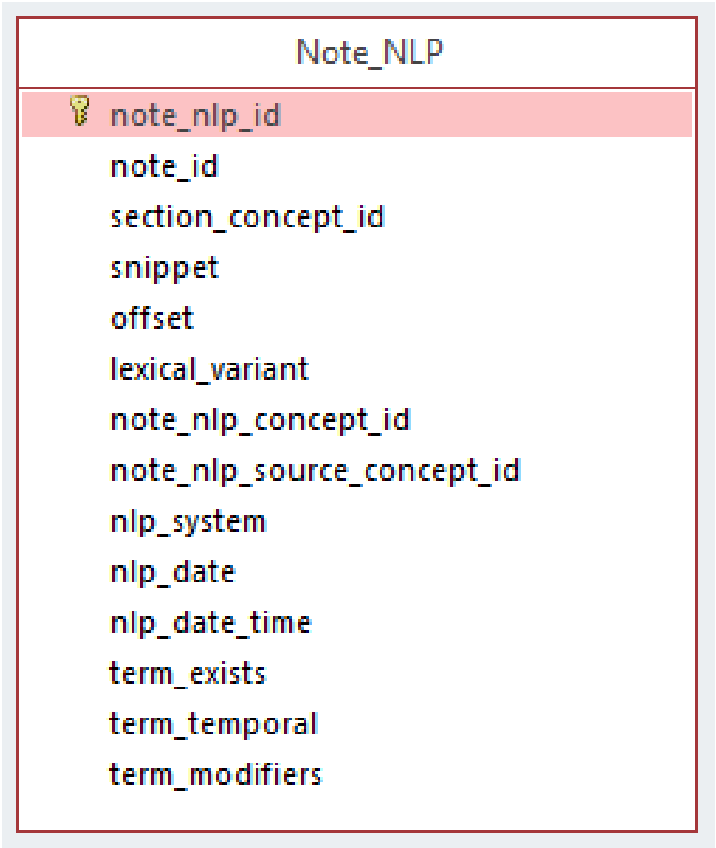
- 收集生物标志物/组织银行

NOTE表



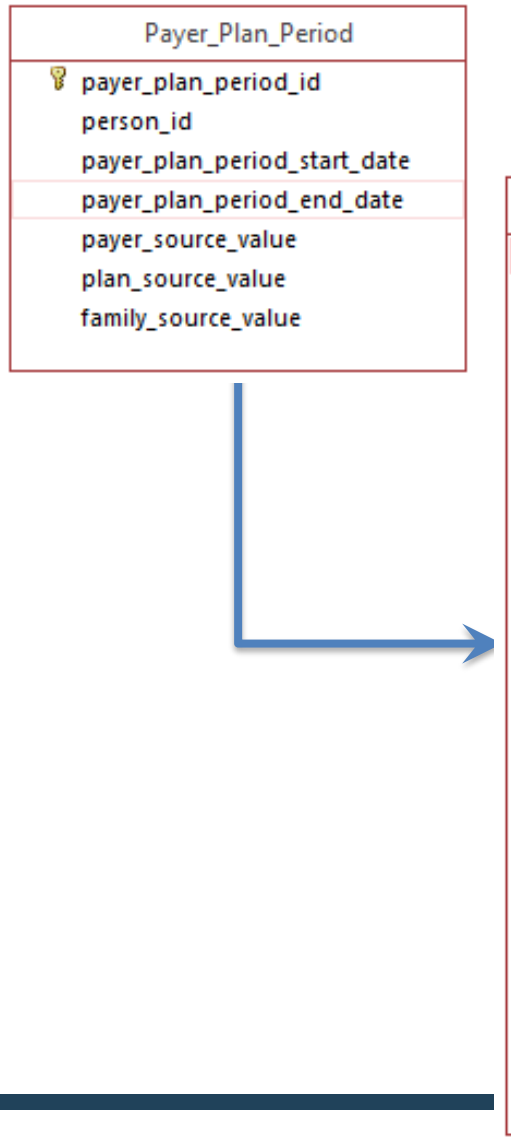
- 记录非结构化的文本

NOTE_NLP表



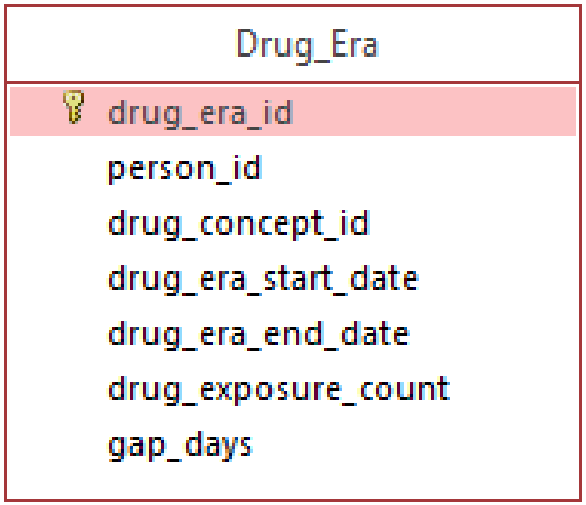
- NOTE_NLP表将对临床病例中的所有NLP输出进行编码。每行代表一份病例中一个识别出来的词

Health Economics



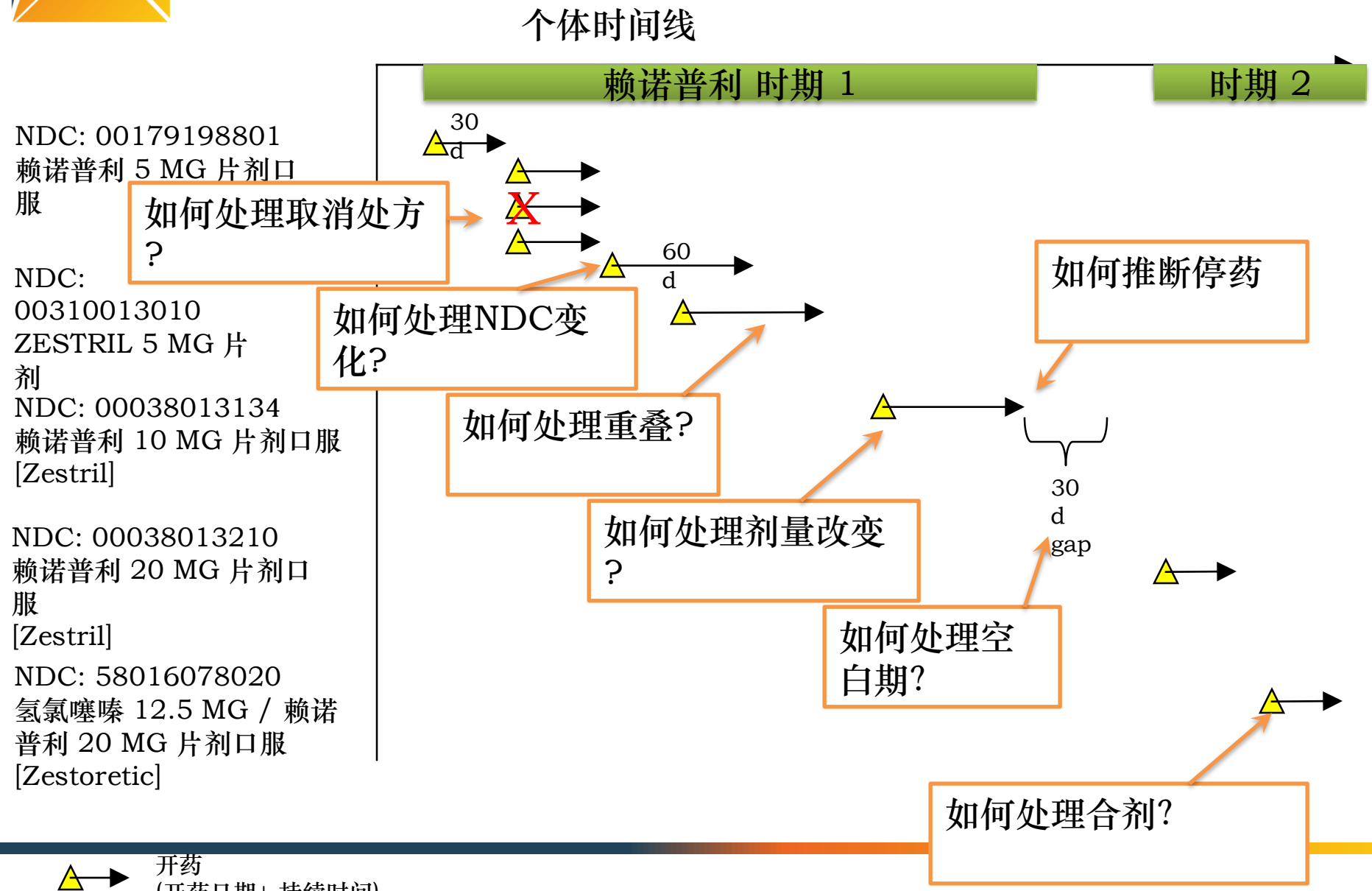
- 所有的费用合并为一个COST 表
- 费用（cost)与相应的观察记录相关联
- 版块由cost_domain_id决定 (e.g. 就诊，疾病，etc.)

DRUG_ERA表



- 对于所有被使用的药物成分的暴露时间的标准化推断
- 通过一些条件从DRUG_EXPOSURE表中的记录运算得出的连续时间段

下图通过一位患者的纵向药房保险数据说明推断的必要性



Some Example Questions

Ex 0

Finding Warfarin

Ex 1

New Users of Warfarin

Ex 2

New Users of Warfarin who are >=65?

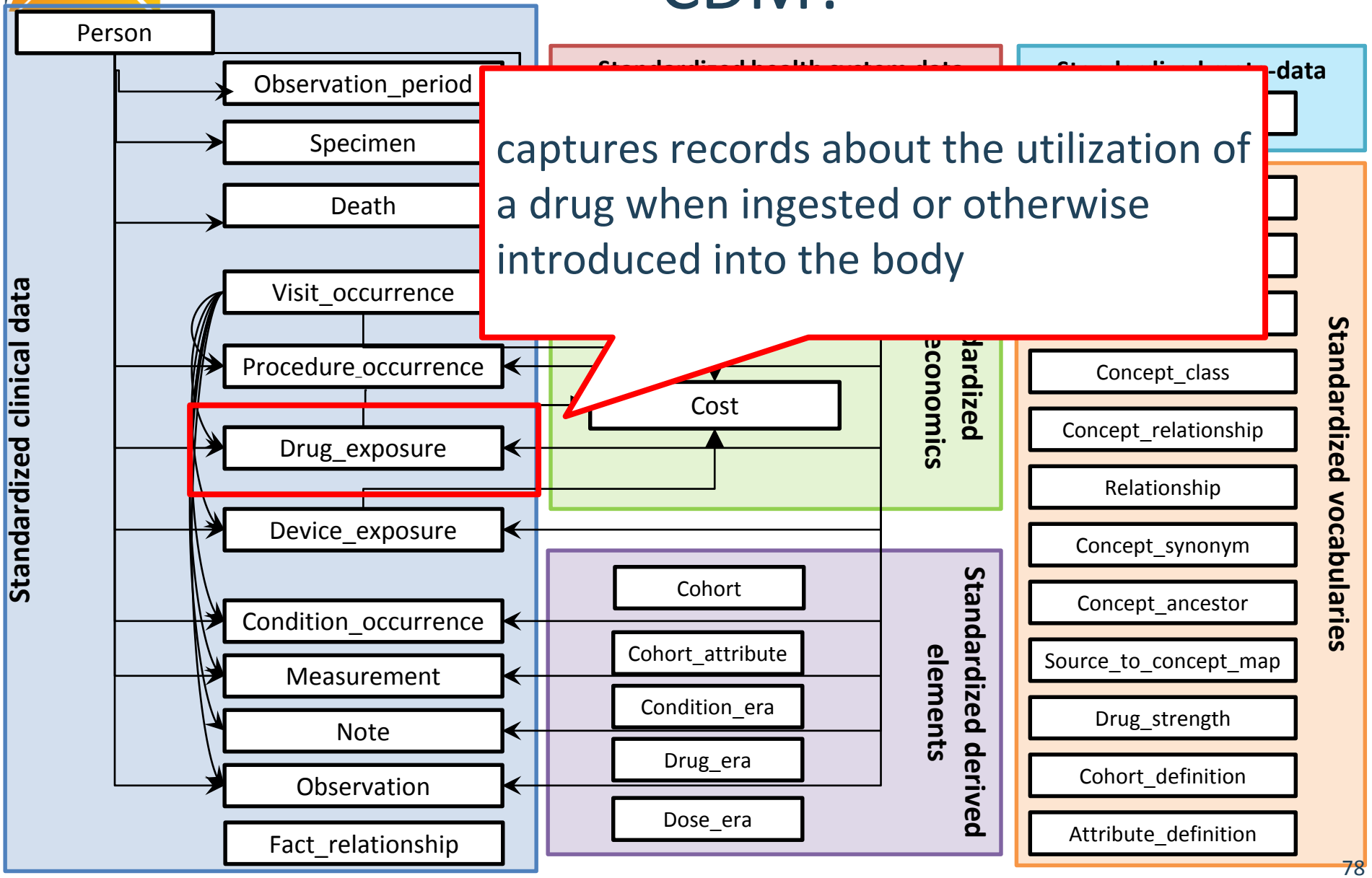
Ex 3

New Users of Warfarin with prior Atrial Fibrillation?

Warfarin Exposure

- Warfarin is a blood thinner that is used to treat/prevent blood clots.
  - Where do you find drug data in the CDM?
  - What codes do I use to define drugs?

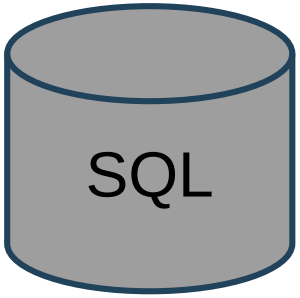
Where are Drug Exposures in the CDM?



How do I define Warfarin?

- When raw data is transformed into the CDM raw source codes are transformed into standard OMOP Vocabulary concepts
- In the CDM, we no longer care what source codes existed in the raw data, we just need to use concept identifiers
- We can use the OMOP Vocabulary to identify all concepts that contain the ingredient warfarin

How do I define Warfarin?



- Writing SQL Statement

- OHDSI Tool ATLAS

Finding Warfarin

Ex 0

```
/*********************************************************************************
*       (Exercise 0) Finding Warfarin
*********************************************************************************/

/*Just looking for the ingredient concept*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE DRUG_CONCEPT_ID = 1310149 /*warfarin*/;
```

0 individuals

```
/*Looking for drugs associated with the ingredient*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR
        WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
);

/*looking for anticoagulants, a class of drugs warfarin belongs*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR
        WHERE ANCESTOR_CONCEPT_ID = 4283987 /*ANTICOAGULANTS (VA Class)*/
);
```

Finding Warfarin

Ex 0

```
/*********************************************************************************
*       (Exercise 0) Finding Warfarin
*********************************************************************************/

/*Just looking for the ingredient concept*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE DRUG_CONCEPT_ID = 1310149 /*warfarin*/;

/*Looking for drugs associated with the ingredient*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR
        WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
);

/*looking for anticoagulants, a class of drugs warfarin belongs*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR
        WHERE ANCESTOR_CONCEPT_ID = 4283987 /*ANTICOAGULANTS (VA Class)*/
);
```

0 individuals

22,247 individuals

Finding Warfarin

Ex 0

```
/*********************************************************************************
*       (Exercise 0) Finding Warfarin
*********************************************************************************/

/*Just looking for the ingredient concept*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE DRUG_CONCEPT_ID = 1310149 /*warfarin*/;

/*Looking for drugs associated with the ingredient*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR
        WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
);

/*looking for anticoagulants, a class of drugs warfarin belongs*/
SELECT COUNT(DISTINCT de.PERSON_ID)
FROM DRUG_EXPOSURE de
WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR
        WHERE ANCESTOR_CONCEPT_ID = 4283987 /*ANTICOAGULANTS (VA Class)*/
);
```

0 individuals

22,247 individuals

32,955 individuals

Some Example Questions

Ex 0 Finding Warfarin

Ex 1 New Users of Warfarin

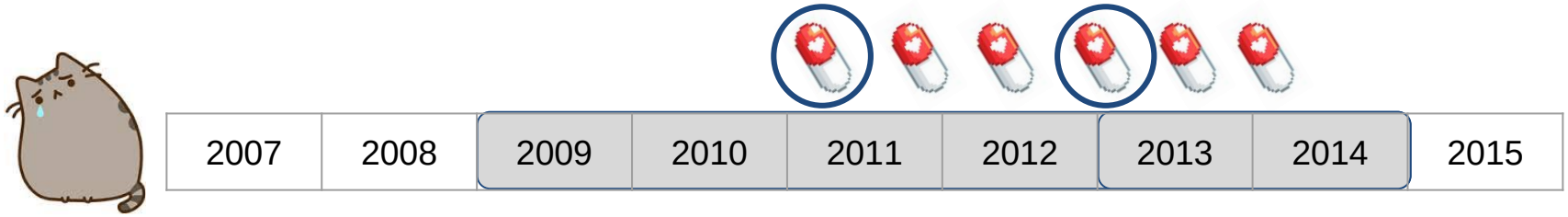
Ex 2 New Users of Warfarin who are >=65?

Ex 3 New Users of Warfarin with prior Atrial Fibrillation?

How do I define new users of a drug?

Ex 1

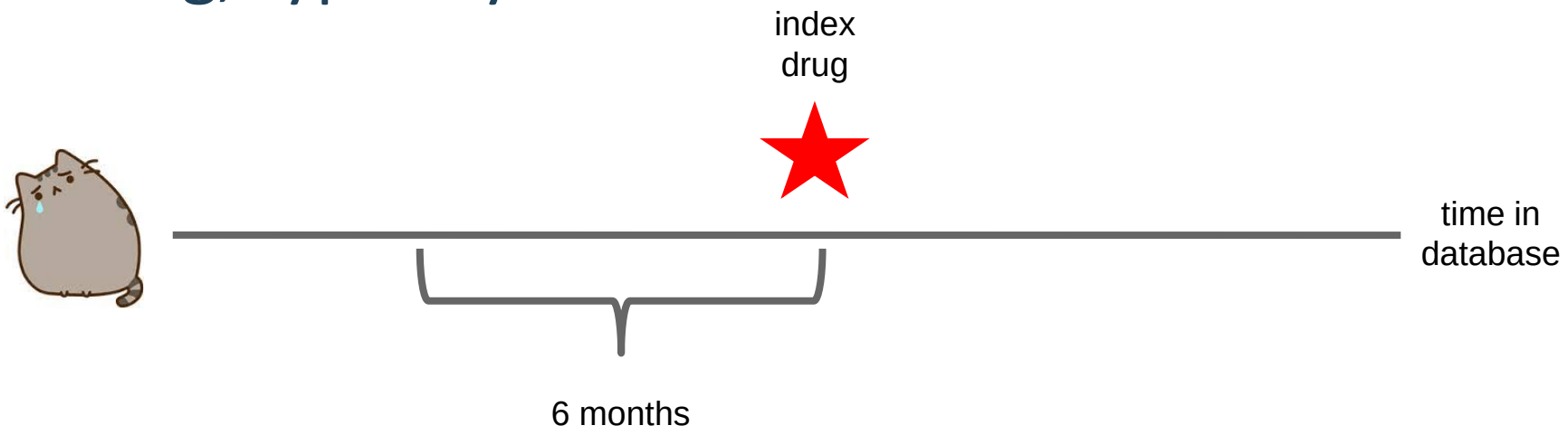
Someone who has recently started taking the drug, typically with a 6 or 12 month wash out



How do I define new users of a drug?

Ex 1

Someone who has recently started taking the drug, typically with a 6 or 12 month wash out



What is Needed in the CDM?

Ex 1

- **OMOP Vocabulary**
  to find the concepts

- **CDM Table DRUG_EXPOSURE**
  to find individuals with exposure

- **CDM Table OBSERVATION_PERIOD**
  to know people's time within the database

New Users of Warfarin

Ex 1

```
/**************************************************************************************
*    (Exercise 1) Warfarin New Users
**************************************************************************************/

WITH CTE_DRUG_INDEX AS (
    SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
    FROM DRUG_EXPOSURE de
    WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
    )
    GROUP BY de.PERSON_ID
)
SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
    (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
FROM CTE_DRUG_INDEX i
    JOIN OBSERVATION_PERIOD op
        ON op.PERSON_ID = i.PERSON_ID
        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
ORDER BY i.PERSON_ID
```

Step 1: Get the codes you need

Ex 1

```
/**********************************************************************************
*    (Exercise 1) Warfarin New Users
**********************************************************************************/

WITH CTE_DRUG_INDEX AS (
    SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
    FROM DRUG_EXPOSURE de
    WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
    )
    GROUP BY de.PERSON_ID
)
SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
    (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
FROM CTE_DRUG_INDEX i
    JOIN OBSERVATION_PERIOD op
        ON op.PERSON_ID = i.PERSON_ID
        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
ORDER BY i.PERSON_ID
```

Step 2: Find Drug Exposures

Ex 1

```
/**********************************************************************************
*    (Exercise 1) Warfarin New Users
**********************************************************************************/

WITH CTE_DRUG_INDEX AS (
    SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
    FROM DRUG_EXPOSURE de
    WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
    )
    GROUP BY de.PERSON_ID
)
SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
    (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
FROM CTE_DRUG_INDEX i
    JOIN OBSERVATION_PERIOD op
        ON op.PERSON_ID = i.PERSON_ID
        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
ORDER BY i.PERSON_ID
```

Step 3: Find New Users

Ex 1

```
/**********************************************************************************
*    (Exercise 1) Warfarin New Users
**********************************************************************************/

WITH CTE_DRUG_INDEX AS (
    SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
    FROM DRUG_EXPOSURE de
    WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
    )
    GROUP BY de.PERSON_ID
)
SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
    (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
FROM CTE_DRUG_INDEX i
    JOIN OBSERVATION_PERIOD op
        ON op.PERSON_ID = i.PERSON_ID
        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
ORDER BY i.PERSON_ID
```

New Users of Warfarin

Ex 1

```
/**************************************************************************************
*    (Exercise 1) Warfarin New Users
**************************************************************************************/

WITH CTE_DRUG_INDEX AS (
    SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
    FROM DRUG_EXPOSURE de
    WHERE de.DRUG_CONCEPT_ID IN (
        SELECT DESCENDANT_CONCEPT_ID
        FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
    )
    GROUP BY de.PERSON_ID
)
SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
    (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
FROM CTE_DRUG_INDEX i
    JOIN OBSERVATION_PERIOD op
        ON op.PERSON_ID = i.PERSON_ID
        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
ORDER BY i.PERSON_ID
```

Some Example Questions

Ex 0 Finding Warfarin

Ex 1 New Users of Warfarin

Ex 2

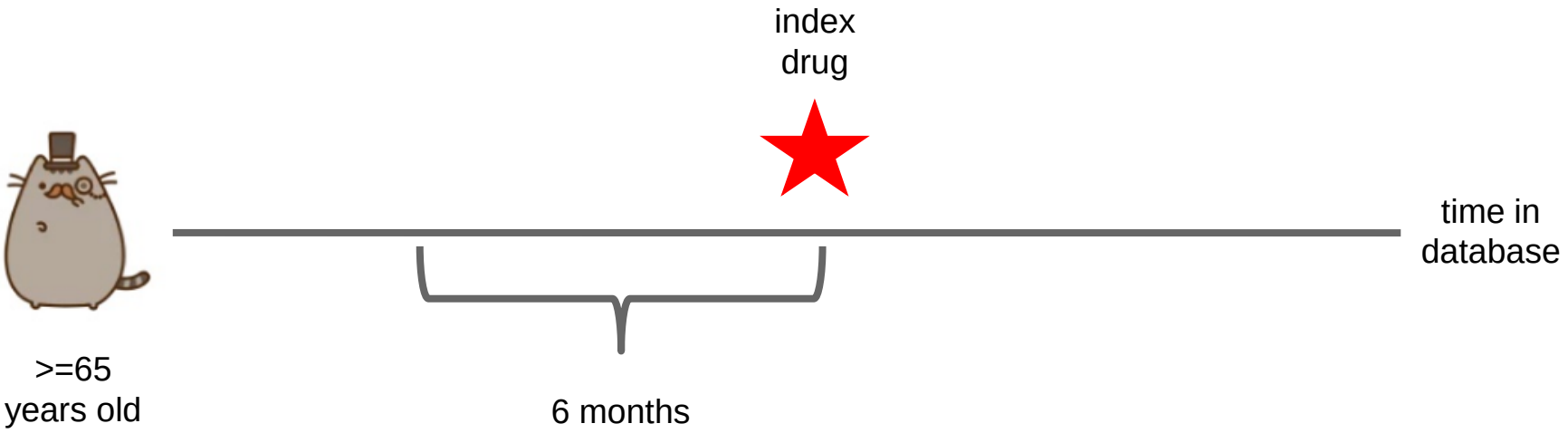
New Users of Warfarin who are >=65?

Ex 3 New Users of Warfarin with prior Atrial Fibrillation?

How do I define new users of warfarin who are >=65?

Ex 2

Someone who has recently started taking the drug, typically with a 6 or 12 month wash out



What is Needed in the CDM?

Ex 2

- **OMOP Vocabulary**
  to find the concepts
- **DRUG_EXPOSURE**
  to find individuals with exposure
- **OBSERVATION_PERIOD**
  to know people's time within the database
- **PERSON**
  to know year of birth

Step 1: Start with the previous query

Ex 2

```
/*********************************************************************************
*       (Exercise 2) Warfarin New Users 65 or Older at Index
*********************************************************************************/

WITH CTE_DRUG_INDEX AS (
        SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
        FROM DRUG_EXPOSURE de
        WHERE de.DRUG_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
        )
        GROUP BY de.PERSON_ID
)
SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
        (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX,
        EXTRACT(YEAR FROM i.INDEX_DATE)-p.YEAR_OF_BIRTH AS AGE_AT_INDEX
FROM CTE_DRUG_INDEX i
        JOIN OBSERVATION_PERIOD op
                ON op.PERSON_ID = i.PERSON_ID
                AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
        JOIN PERSON p
                ON p.PERSON_ID = i.PERSON_ID
WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
AND EXTRACT(YEAR FROM i.INDEX_DATE)-p.YEAR_OF_BIRTH >= 65
ORDER BY i.PERSON_ID
```

Step 2: Add the Person Table to calculate age

Ex 2

```
/*********************************************************************************
*       (Exercise 2) Warfarin New Users 65 or Older at Index
*********************************************************************************/

WITH CTE_DRUG_INDEX AS (
        SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
        FROM DRUG_EXPOSURE de
        WHERE de.DRUG_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
        )
        GROUP BY de.PERSON_ID
)
SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
        (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX,
        EXTRACT(YEAR FROM i.INDEX_DATE)-p.YEAR_OF_BIRTH AS AGE_AT_INDEX
FROM CTE_DRUG_INDEX i
        JOIN OBSERVATION_PERIOD op
                ON op.PERSON_ID = i.PERSON_ID
                AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
        JOIN PERSON p
                ON p.PERSON_ID = i.PERSON_ID
WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
AND EXTRACT(YEAR FROM i.INDEX_DATE)-p.YEAR_OF_BIRTH >= 65
ORDER BY i.PERSON_ID
```

New Users of Warfarin >= 65 years of age

Ex 2

Try running this on your own!

```
/**********************************************************************************
*       (Exercise 2) Warfarin New Users 65 or Older at Index
**********************************************************************************/

WITH CTE_DRUG_INDEX AS (
        SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
        FROM DRUG_EXPOSURE de
        WHERE de.DRUG_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
        )
        GROUP BY de.PERSON_ID
)
SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
        (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX,
        EXTRACT(YEAR FROM i.INDEX_DATE)-p.YEAR_OF_BIRTH AS AGE_AT_INDEX
FROM CTE_DRUG_INDEX i
        JOIN OBSERVATION_PERIOD op
                ON op.PERSON_ID = i.PERSON_ID
                AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
        JOIN PERSON p
                ON p.PERSON_ID = i.PERSON_ID
WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
AND EXTRACT(YEAR FROM i.INDEX_DATE)-p.YEAR_OF_BIRTH >= 65
ORDER BY i.PERSON_ID
```

How many people do you get?

Some Example Questions

Ex 0 — Finding Warfarin

Ex 1 — New Users of Warfarin

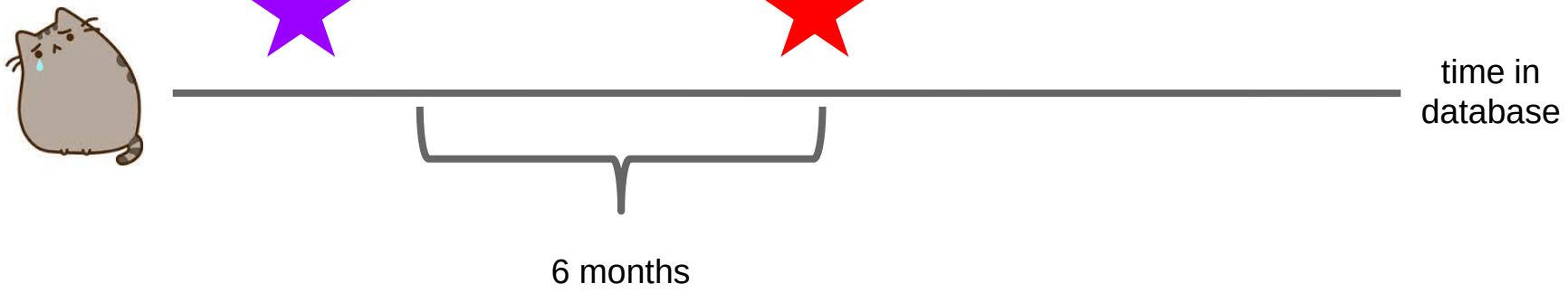
Ex 2 — New Users of Warfarin who are >=65?

Ex 3

New Users of Warfarin with prior Atrial Fibrillation?

How do I define new users of Warfarin with prior Atrial Fibrillation?

Ex 3



What is Needed in the CDM?

Ex 3

- **OMOP Vocabulary**
  to find the concepts
- **DRUG_EXPOSURE**
  to find individuals with exposure
- **OBSERVATION_PERIOD**
  to know people's time within the database
- **PERSON**
  to know year of birth
- **CONDITION_OCCURRENCE**
  to find presence of a disease

Step 1: Start with the Ex 1 query

Ex 3

```
/**********************************************************************************
*       (Exercise 3) Warfarin New Users With Prior AFIB
**********************************************************************************/

WITH CTE_DRUG_INDEX AS (
        SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
        FROM DRUG_EXPOSURE de
        WHERE de.DRUG_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
        )
        GROUP BY de.PERSON_ID
),
CTE_DRUG_NEW_USERS AS (
        SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
                (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
        FROM CTE_DRUG_INDEX i
                JOIN OBSERVATION_PERIOD op
                        ON op.PERSON_ID = i.PERSON_ID
                        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
        WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
)
SELECT nu.*, MIN(nu.INDEX_DATE-co.CONDITION_START_DATE) AS DAYS_OF_CLOSEST_AFIB_PRIOR_TO_INDEX
FROM CTE_DRUG_NEW_USERS nu
        JOIN CONDITION_OCCURRENCE co
                ON co.PERSON_ID = nu.PERSON_ID
                AND co.CONDITION_START_DATE BETWEEN nu.OBSERVATION_PERIOD_START_DATE AND nu.OBSERVATION_PERIOD_END_DATE
WHERE co.CONDITION_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID =  313217 /*Atrial fibrillation*/
)
AND co.CONDITION_START_DATE < nu.INDEX_DATE
GROUP BY nu.PERSON_ID, nu.INDEX_DATE, nu.OBSERVATION_PERIOD_START_DATE, nu.OBSERVATION_PERIOD_END_DATE, nu.DAYS_BEFORE_INDEX
ORDER BY nu.PERSON_ID
```

Step 2: Define Atrial Fibrillation

Ex 3

```
/**********************************************************************************
*       (Exercise 3) Warfarin New Users With Prior AFIB
**********************************************************************************/

WITH CTE_DRUG_INDEX AS (
        SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
        FROM DRUG_EXPOSURE de
        WHERE de.DRUG_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
        )
        GROUP BY de.PERSON_ID
),
CTE_DRUG_NEW_USERS AS (
        SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
                (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
        FROM CTE_DRUG_INDEX i
                JOIN OBSERVATION_PERIOD op
                        ON op.PERSON_ID = i.PERSON_ID
                        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
        WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
)
SELECT nu.*, MIN(nu.INDEX_DATE-co.CONDITION_START_DATE) AS DAYS_OF_CLOSEST_AFIB_PRIOR_TO_INDEX
FROM CTE_DRUG_NEW_USERS nu
        JOIN CONDITION_OCCURRENCE co
                ON co.PERSON_ID = nu.PERSON_ID
                AND co.CONDITION_START_DATE BETWEEN nu.OBSERVATION_PERIOD_START_DATE AND nu.OBSERVATION_PERIOD_END_DATE
WHERE co.CONDITION_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID =  313217 /*Atrial fibrillation*/
)
AND co.CONDITION_START_DATE < nu.INDEX_DATE
GROUP BY nu.PERSON_ID, nu.INDEX_DATE, nu.OBSERVATION_PERIOD_START_DATE, nu.OBSERVATION_PERIOD_END_DATE, nu.DAYS_BEFORE_INDEX
ORDER BY nu.PERSON_ID
```

Step 3: Prior Atrial Fibrillation

Ex 3

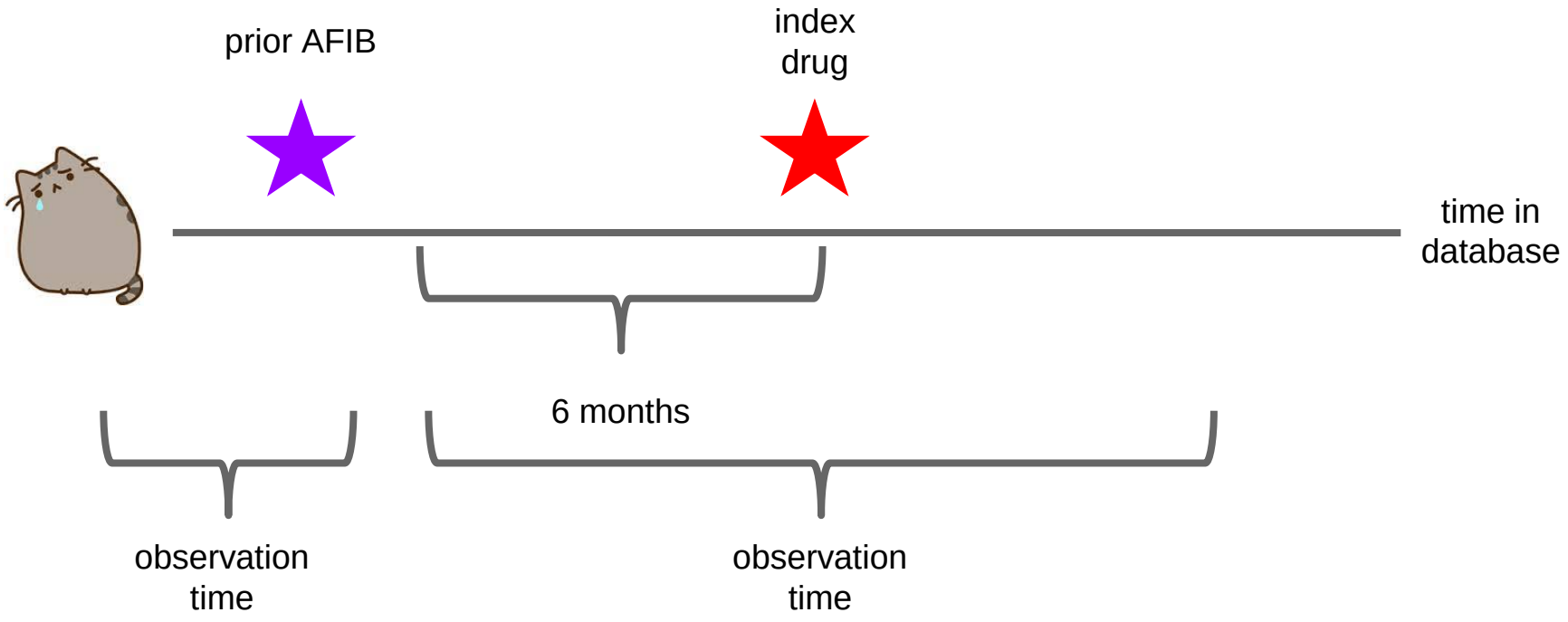
```
/**********************************************************************************
*       (Exercise 3) Warfarin New Users With Prior AFIB
**********************************************************************************/

WITH CTE_DRUG_INDEX AS (
        SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
        FROM DRUG_EXPOSURE de
        WHERE de.DRUG_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
        )
        GROUP BY de.PERSON_ID
),
CTE_DRUG_NEW_USERS AS (
        SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
                (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
        FROM CTE_DRUG_INDEX i
                JOIN OBSERVATION_PERIOD op
                        ON op.PERSON_ID = i.PERSON_ID
                        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND
        WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
)
SELECT nu.*, MIN(nu.INDEX_DATE-co.CONDITION_START_DATE) AS DAYS_OF_CLOSEST_AFIB_PRIOR
FROM CTE_DRUG_NEW_USERS nu
        JOIN CONDITION_OCCURRENCE co
                ON co.PERSON_ID = nu.PERSON_ID
                AND co.CONDITION_START_DATE BETWEEN nu.OBSERVATION_PERIOD_START_DATE AND nu.OBSERVATION_PERIOD_END_DATE
WHERE co.CONDITION_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID =  313217 /*Atrial fibrillation*/
)
AND co.CONDITION_START_DATE < nu.INDEX_DATE
GROUP BY nu.PERSON_ID, nu.INDEX_DATE, nu.OBSERVATION_PERIOD_START_DATE, nu.OBSERVATION_PERIOD_END_DATE, nu.DAYS_BEFORE_INDEX
ORDER BY nu.PERSON_ID
```

Keeps condition within the same observable time, exclude if you want all time prior

How do I define new users of Warfarin with prior Atrial Fibrillation?

Ex 3



New Users of Warfarin with prior Atrial Fibrillation

Ex 3

```
/**********************************************************************************
*       (Exercise 3) Warfarin New Users With Prior AFIB
**********************************************************************************/

WITH CTE_DRUG_INDEX AS (
        SELECT de.PERSON_ID, MIN(de.DRUG_EXPOSURE_START_DATE) AS INDEX_DATE
        FROM DRUG_EXPOSURE de
        WHERE de.DRUG_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID = 1310149 /*warfarin*/
        )
        GROUP BY de.PERSON_ID
),
CTE_DRUG_NEW_USERS AS (
        SELECT i.PERSON_ID, i.INDEX_DATE, op.OBSERVATION_PERIOD_START_DATE, op.OBSERVATION_PERIOD_END_DATE,
                (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) AS DAYS_BEFORE_INDEX
        FROM CTE_DRUG_INDEX i
                JOIN OBSERVATION_PERIOD op
                        ON op.PERSON_ID = i.PERSON_ID
                        AND i.INDEX_DATE BETWEEN op.OBSERVATION_PERIOD_START_DATE AND op.OBSERVATION_PERIOD_END_DATE
        WHERE (i.INDEX_DATE-op.OBSERVATION_PERIOD_START_DATE) >= 180
)
SELECT nu.*, MIN(nu.INDEX_DATE-co.CONDITION_START_DATE) AS DAYS_OF_CLOSEST_AFIB_PRIOR_TO_INDEX
FROM CTE_DRUG_NEW_USERS nu
        JOIN CONDITION_OCCURRENCE co
                ON co.PERSON_ID = nu.PERSON_ID
                AND co.CONDITION_START_DATE BETWEEN nu.OBSERVATION_PERIOD_START_DATE AND nu.OBSERVATION_PERIOD_END_DATE
WHERE co.CONDITION_CONCEPT_ID IN (
                SELECT DESCENDANT_CONCEPT_ID FROM CONCEPT_ANCESTOR WHERE ANCESTOR_CONCEPT_ID =  313217 /*Atrial fibrillation*/
)
AND co.CONDITION_START_DATE < nu.INDEX_DATE
GROUP BY nu.PERSON_ID, nu.INDEX_DATE, nu.OBSERVATION_PERIOD_START_DATE, nu.OBSERVATION_PERIOD_END_DATE, nu.DAYS_BEFORE_INDEX
ORDER BY nu.PERSON_ID
```

Extract-Transform-Load (ETL)

ETL: 真实案例

PharMetrics Plus

CLAIMS

pat_id	claimno	from_dt	to_dt	diagprc_ind	Diag_admit	diag1
05917921689	IPA333393946	1/5/2006	1/5/2006	1	41071	41071

LRx/Dx

MEDICAL_CLAIMS

md_clm_id	ims_pat_nbr	dt_of_service	rxer_id	diag_cd
95963982102	80445908	8/1/2012 0:00	680488	41071

German DA

Problem Events

db_country	international_ practice_num	international_ doctor_num	international_ patient_num	age
GE	GE6326	GE8784	GE46478747	

Diagnosis

db_country	international_dia gnosis_num	diagnosis_num	icd10_4_(
GE	GE2397573	2397573	I21.4

Ambulatory EMR

Problem

Patient_id_synth	Diag_dt	Icd10_cd
271138	4/11/2013	I214

4个真实的观察数据库，所有数据库均包含诊断为“急性心内膜下梗死”的患者的住院病例，

- 没有一个表名称相同…
- 没有一个变量名称相同…。
- 不同的表结构（行与列）
- 不同的表示方法（带和不带小数点）
- 不同的编码方案（ICD9与ICD10）

ETL 对 OMOP CDM意味着什么？
标准化结构和内容

PharMetrics Plus

Inpatient Claims

pat_id	claimno	from_dt	to_dt	diagprc_ind	Diag_admit
05917921689	IPA333393946	1/5/2006	1/5/2006	1	41071

针对临床诊断，人群水平的估计和患者水平的预测的大规模分析优化了结构

PharMetrics Plus

CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID
05917921689	1/5/2006	41071	Inpatient claims - primary position

应用国际化术语标准的内容可以被应用于任何数据库

PharMetrics Plus

CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION _TYPE_CONCEPT_ID	CONDITION_SOURCE_CONCEPT_ID	CONDITION_CONCEPT_ID
05917921689	1/5/2006	41071	Inpatient claims - primary position	44825429	444406

OMOP CDM = 标准化的结构：相同的表，相同的字段，相同的数据类型，不同来源数据的相同表现方式

PharMetrics Plus

CLAIMS

pat_id	claimno	from_dt	to_dt	diagprc_ind	Diag_admit	diag1
05917921689	IPA333393946	1/5/2006	1/5/2006	1	41071	41071

LRx/Dx

MEDICAL_CLAIMS

md_clm_id	ims_pat_nbr	dt_of_service	rxer_id	diag_cd
95963982102	80445908	8/1/2012 0:00	680488	41071

German DA

Problem Events

db_country	international_ practice_num	international_ doctor_num	international_ patient_num	age_at_event	date_of_event	international_ diagnosis_nu m
GE	GE6326	GE8784	GE46478747	20	11/19/2014 0:00	GE2397573

Diagnosis

db_country	international_dia gnosis_num	diagnosis_num	icd10_4_code	icd10_3_text	diagnosis_conf idence
GE	GE2397573	2397573	I21.4	Non-ST elevation (NSTEMI) myocardial infarction	Confirmed

Ambulatory EMR

Problem

Patient_id_synth	Diag_dt	icd10_cd
271138	4/11/2013	I214

- 为大规模分析优化的一致结构
- 该结构保留所有来源内容和出处

PharMetrics Plus: CONDITION_OCCURRENCE

PERSON_ID	CONDITION_ START_DATE	CONDITION _SOURCE_V ALUE	CONDITION_TYPE_CONCEPT_ID
157033702	1/5/2006	41071	Inpatient claims - primary position

LRX/DX: CONDITION_OCCURRENCE

PERSON_ID	CONDITION_ START_DATE	CONDITION _SOURCE_V ALUE	CONDITION_TYPE_CONCEPT_ID
80445908	8/1/2012	41071	Primary Condition

German DA : CONDITION_OCCURRENCE

PERSON_ID	CONDITION_ START_DATE	CONDITION _SOURCE_V ALUE	CONDITION_TYPE_CONCEPT_ID
46478747	11/19/2014	I21.4	EHR problem list entry

Ambulatory EMR : CONDITION_OCCURRENCE

PERSON_ID	CONDITION_ START_DATE	CONDITION _SOURCE_V ALUE	CONDITION_TYPE_CONCEPT_ID
271138	4/11/2013	I214	Primary Condition

OMOP CDM = 标准化内容: 整合不同来源的通用术语集

PharMetrics Plus: CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID	CONDITION_SOURCE_CONCEPT_ID	CONDITION_CONCEPT_ID
05917921689	1/5/2006	41071	Inpatient claims - primary position	4482542	444406

LRx/Dx: CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID	CONDITION_SOURCE_CONCEPT_ID	CONDITION_CONCEPT_ID
80445908	8/1/2012	41071	Primary Condition	4482542	444406

German DA : CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID	CONDITION_SOURCE_CONCEPT_ID	CONDITION_CONCEPT_ID
6478747	11/19/2014	I21.4	EHR problem list entry	4557208	444406

Ambulatory EMR : CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID	CONDITION_SOURCE_CONCEPT_ID	CONDITION_CONCEPT_ID
271138	4/11/2013	I214	Primary Condition	4557208	444406

- 将词汇标准化为共同参照标准（ICD9 / 10→SNOMED）
- 将源代码映射到每个模块中，语言不通亦可交流

- 标准化所有术语集中唯一的源代码
- 不用担心格式问题或代码重复

ETL 流程：角色

- 团队成员
  - CDM 专家
  - 本地数据专家
  - 数据工程师
  - 医学专业人员
  - 商业利益相关方





标准

- Patients without transaction 无记录病人
- 数据整理
  - 病人ID 重复
  - 无效代码记录 (e.g. ‘000’)
- 如何处理吸烟信息

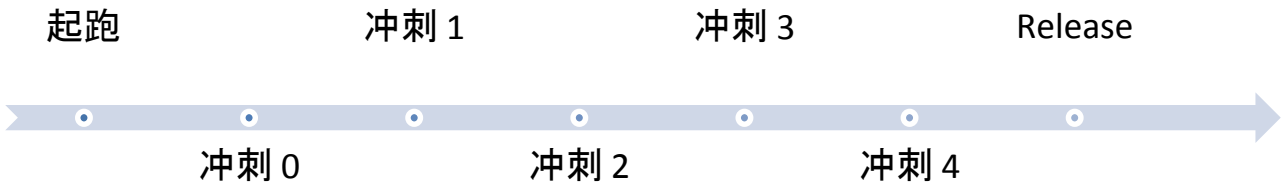
THEMIS

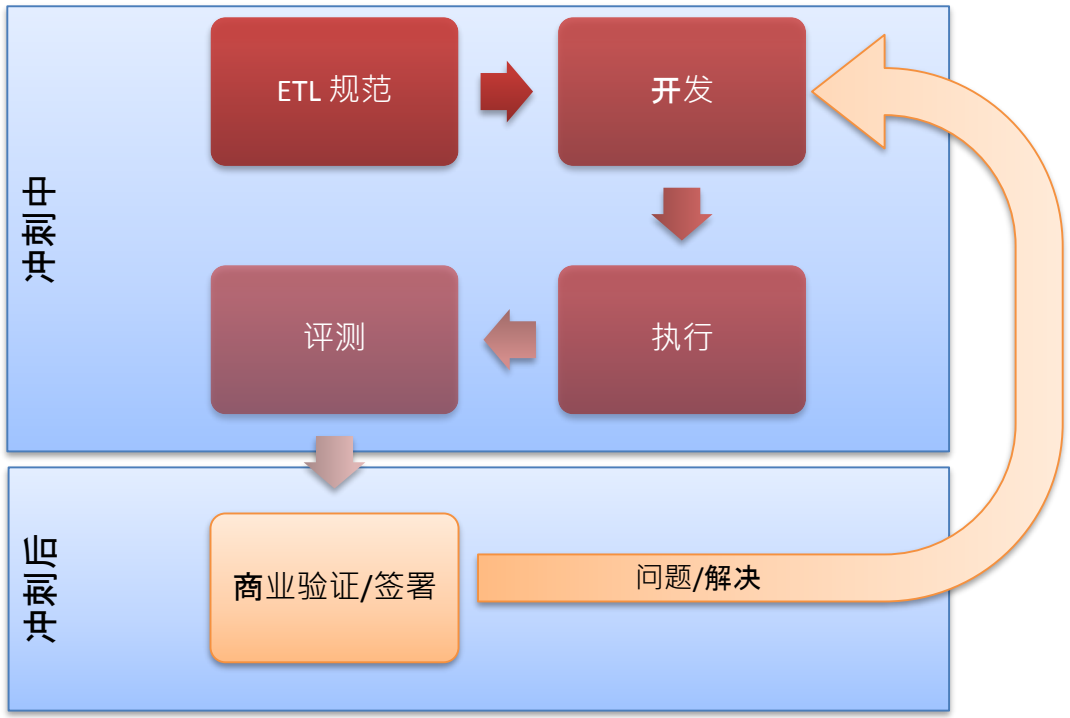
https://github.com/OHDSI/CommonDataModel/wiki

CDM 版本控制

- 工作组一个月开一次会以讨论CDM的修改方案
- 所有CDM的文档，版本，提案均保持于 Github
  - –https://github.com/OHDSI/CommonDataModel
  - –提案会被作为Github的问题进行追踪和讨论
- 工作组会议信息可在 wiki page查询
- 更多信息请联系 Clair Blacketer (mblacke@its.jnj.com)

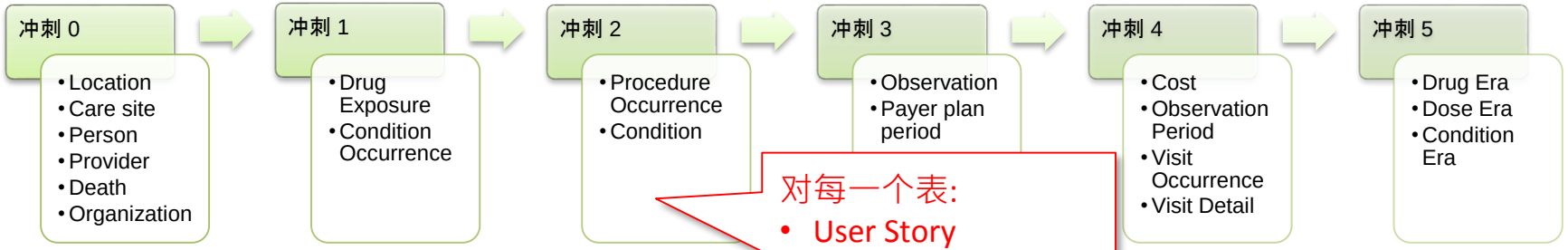
ETL 流程：敏捷Agile





OHDSI ETL 流程举例

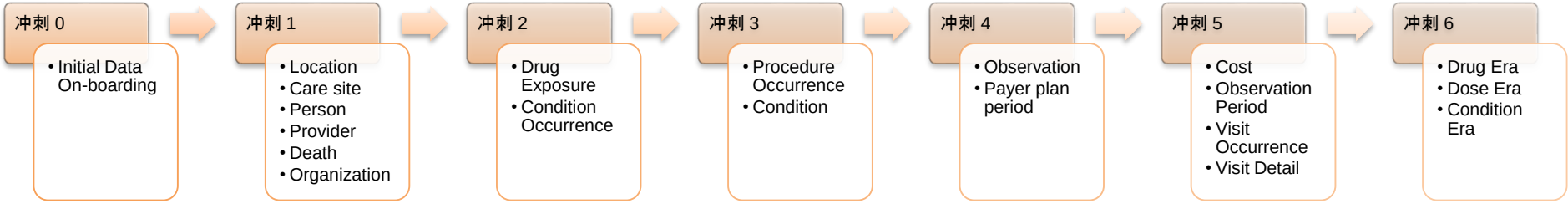
分析 – 创建ETL规范/故事



对每一个表:

- User Story
- White Rabbit
- Vocabulary Mapping
- ETL specs

开发 – 应用/验证

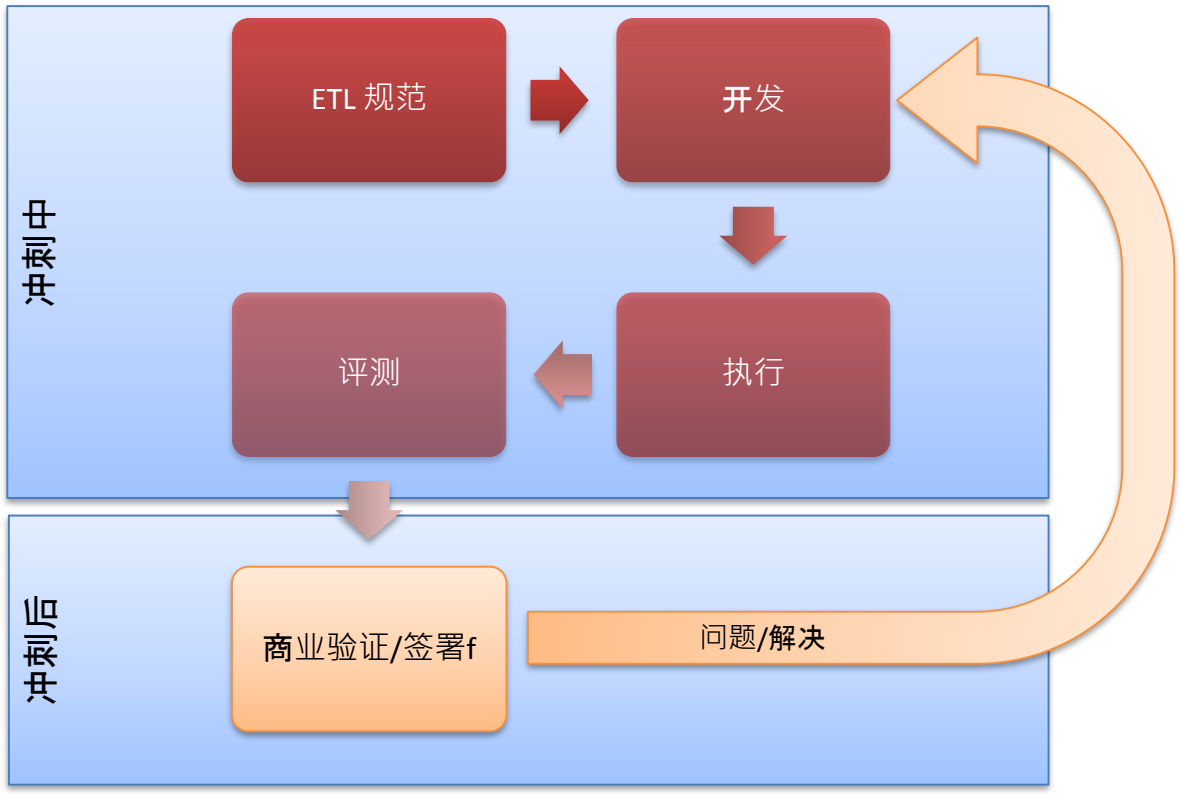


OHDSI ETL资源

Rabbits（兔子）

Usagi（兔子-日语）

Atlas 数据源
(Achilles)



应用

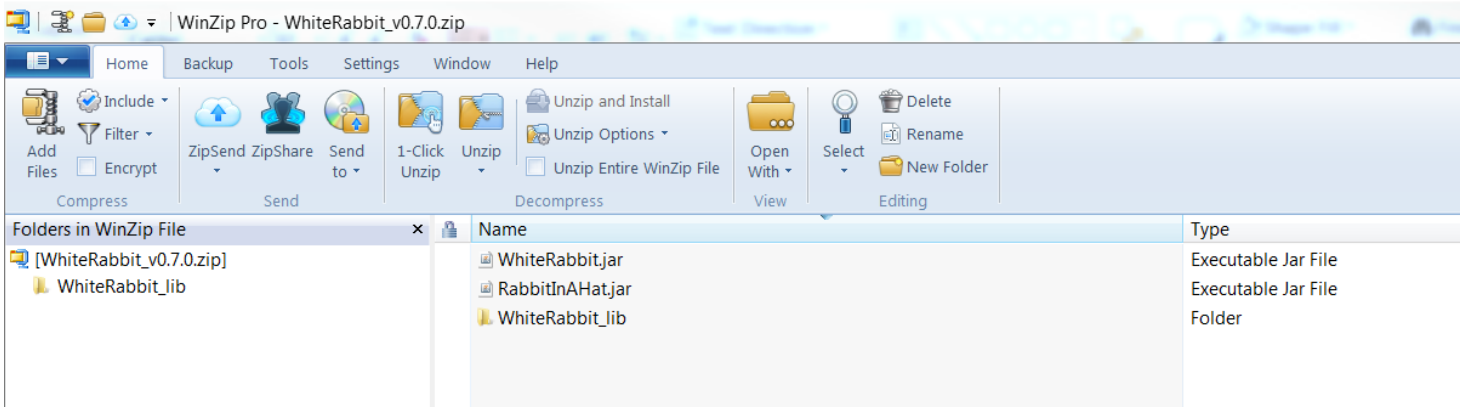
使用 WhiteRabbit 和 Rabbit-In-A-Hat 来搭建一个ETL

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Github

https://github.com/OHDSI/WhiteRabbit

- 要求 Java 1.7 及以上
- 下载最新版White Rabbit 压缩文件



White Rabbit

- 分析数据库的结构和内容
- 扫描全部表格和区域
- 生成扫描报告:
  - 表格，区域内的信息，以及值的频率分布
- 设定值的最低频率下限以保护病人隐私
- 设定最大行数上限以提高效率
- 支持 SQL Server, Oracle, PostgreSQL, MySQL, 和 CSV 文件
- 开源Java App

扫描结果阅读

- XLSX 文件内的一系列选项卡（tabs）

  - Overview Tab
    分析的各个表格的定义，该选项卡具有唯一性

  - Table Tab(s)
    每列的摘要列，会有与选择要分析的表一样多的选项卡

Overview Tab

- 定义扫描的表

	A	B	C	D	E	F	G	H
1	Table	Field	Type	Max lengt	N rows	N rows ch	Fraction empty	
2	Beneficiary_Summary.csv	DESYNPUF	varchar	16	-1	99999	0	
3	Beneficiary_Summary.csv	BENE_BIR	int	8	-1	99999	0	
4	Beneficiary_Summary.csv	BENE_DEA	int	8	-1	99999	0.98449	
5	Beneficiary_Summary.csv	BENE_SEX	int	1	-1	99999	0	
6	Beneficiary_Summary.csv	BENE_RAC	int	1	-1	99999	0	
7	Beneficiary_Summary.csv	BENE_ESR	varchar	1	-1	99999	0	
8	Beneficiary_Summary.csv	SP_STATE_	int	2	-1	99999	0	
9	Beneficiary_Summary.csv	BENE_COU	int	3	-1	99999	0	
10	Beneficiary_Summary.csv	BENE_HI_(	int	2	-1	99999	0	
11	Beneficiary_Summary.csv	BENE_SMI	int	2	-1	99999	0	
12	Beneficiary_Summary.csv	BENE_HM	int	2	-1	99999	0	
13	Beneficiary_Summary.csv	PLAN_CVF	int	2	-1	99999	0	
14	Beneficiary_Summary.csv	SP_ALZHD	int	1	-1	99999	0	
15	Beneficiary_Summary.csv	SP_CHF	int	1	-1	99999	0	
16	Beneficiary_Summary.csv	SP_CHRNI	int	1	-1	99999	0	
17	Beneficiary_Summary.csv	SP_CNCR	int	1	-1	99999	0	
18	Beneficiary_Summary.csv	SP_COPD	int	1	-1	99999	0	
19	Beneficiary_Summary.csv	SP_DEPRE	int	1	-1	99999	0	
20	Beneficiary_Summary.csv	SP_DIABE	int	1	-1	99999	0	
21	Beneficiary_Summary.csv	SP_ISCHM	int	1	-1	99999	0	

Table Tabs

- Beneficiary_Summary.csv 表的定义。每条记录代表一个模拟medicare受保人

#	Variable names	Labels
1	*DESYNPUF_ID*	DESYNPUF: Beneficiary Code
2	*BENE_BIRTH_DT*	DESYNPUF: Date of birth
3	*BENE_DEATH_DT*	DESYNPUF: Date of death
4	*BENE_SEX_IDENT_CD*	DESYNPUF: Sex
5	*BENE_RACE_CD*	DESYNPUF: Beneficiary Race Code

Variable Name: BENE_BIRTH_DT

Type: Num

Format: YYYYMMDD

	A	B	C	D	E	F	G	H
1	DESYNPUF_ID	Frequency	BENE_BIRTH_DT	Frequency	BENE_DEATH_DT	Frequency	BENE_SEX_IDENT_CD	Frequency
2	List truncated...		19421001	466		98448	2	55211
3			19410401	434	20080901	151	1	44788
4			19390401	433	20080101	142		
5			19431101	433	20081001	137		
6			19400301	428	20080301	135		
7			19410501	426	20081101	134		
8			19390301	414	20080401	131		
9			19400501	414	20080701	129		
10			19410801	414	20080501	125		
11			19391201	413	20080801	120		
12			19431001	413	20081201	119		
13			19411001	412	20080201	118		
				412	20080601	110		
				409				

Beneficiary_Summary.csv

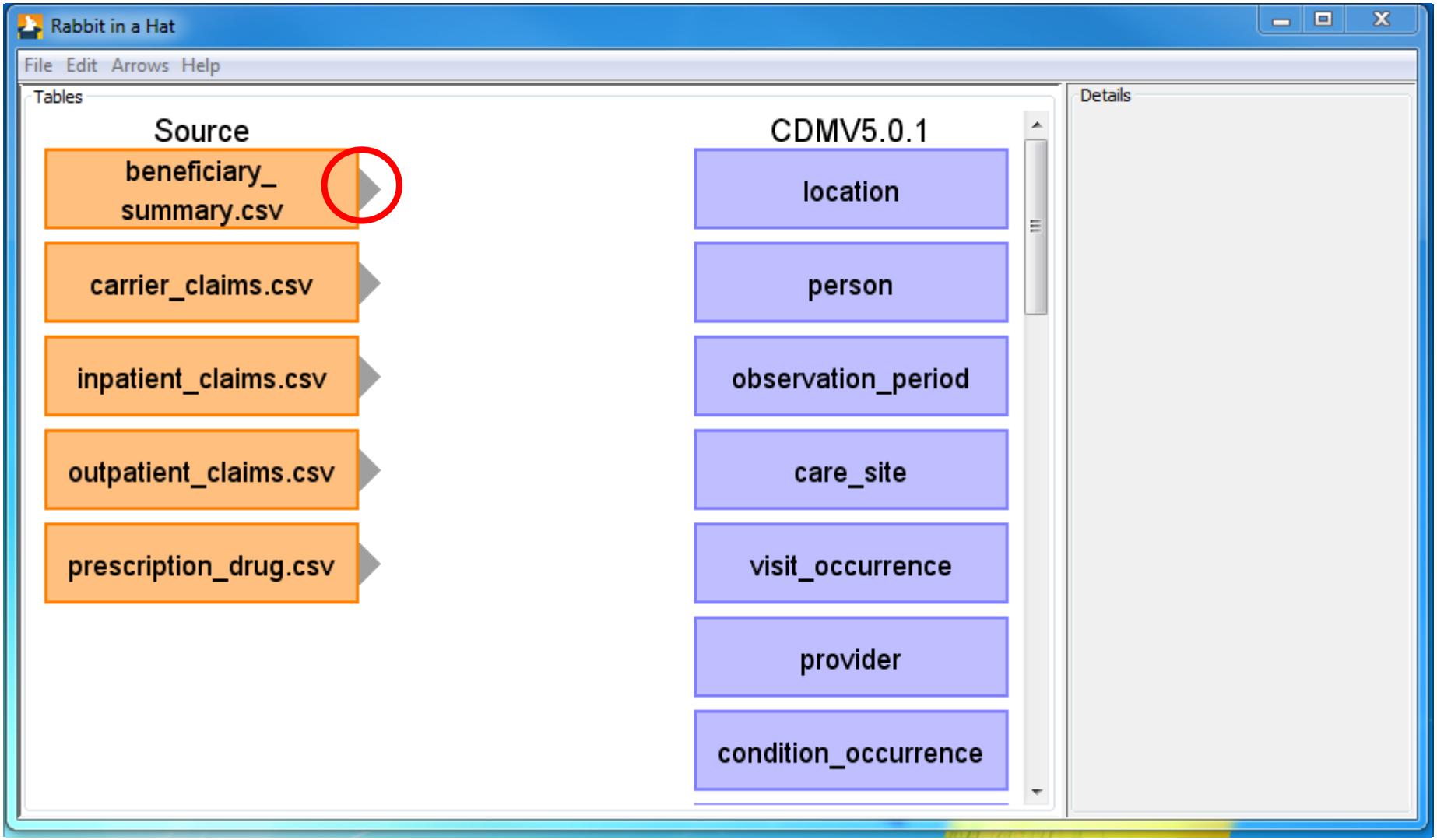
Rabbit In a Hat

- 已整合进WhiteRabbit下载文件
- 使用WhiteRabbit的信息来帮助您生成ETL的流程文档
- 帮您保持设计的逻辑一致性，但不会生成ETL代码

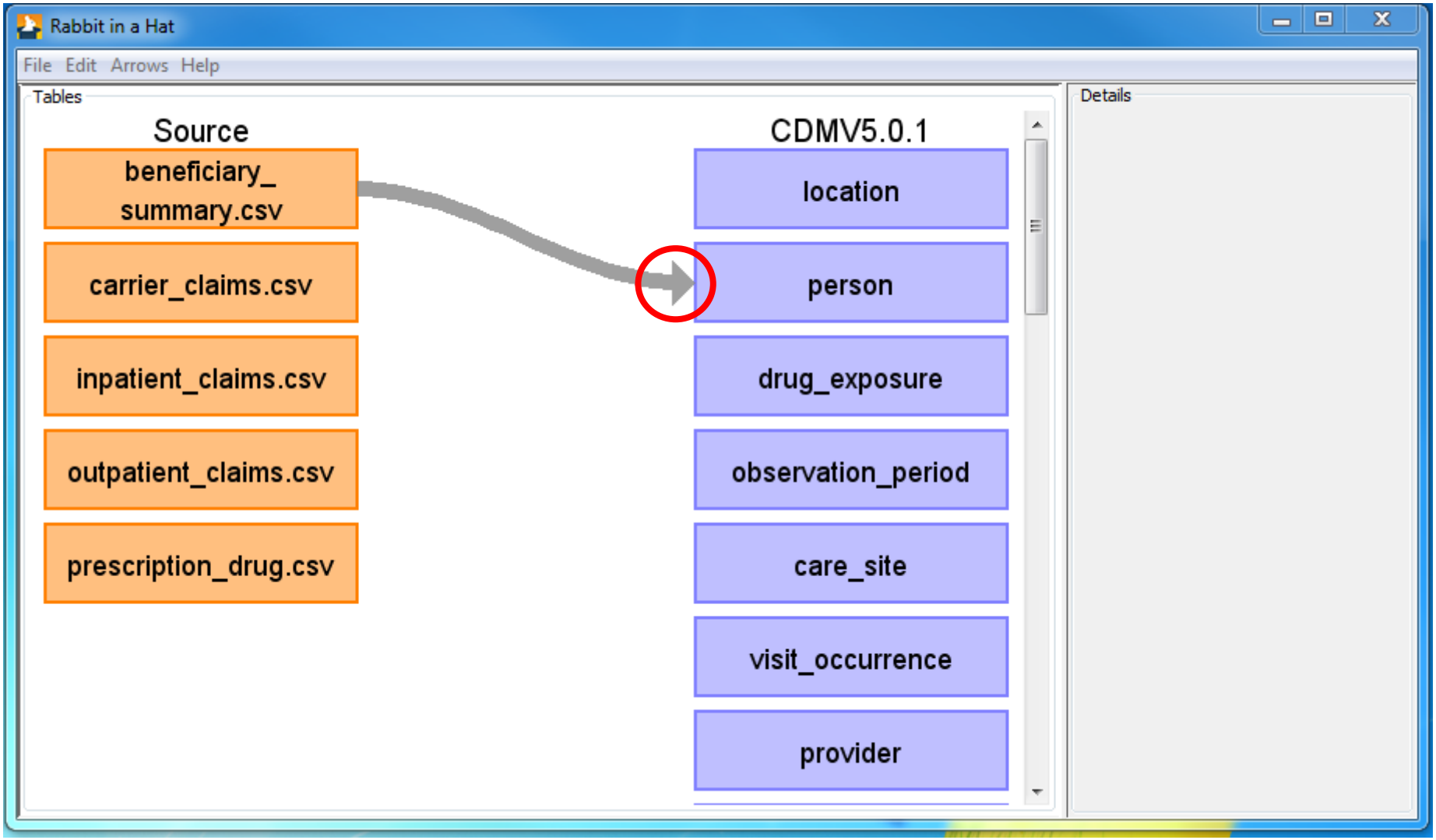
开发ETL的流程

- 召集适合的人
- 需要一段集中工作时间
- 把所有的原始数据表映射到CDM表
- 回顾每个原始数据表映射至CDM表的具体信息
- 生成最终ETL文档

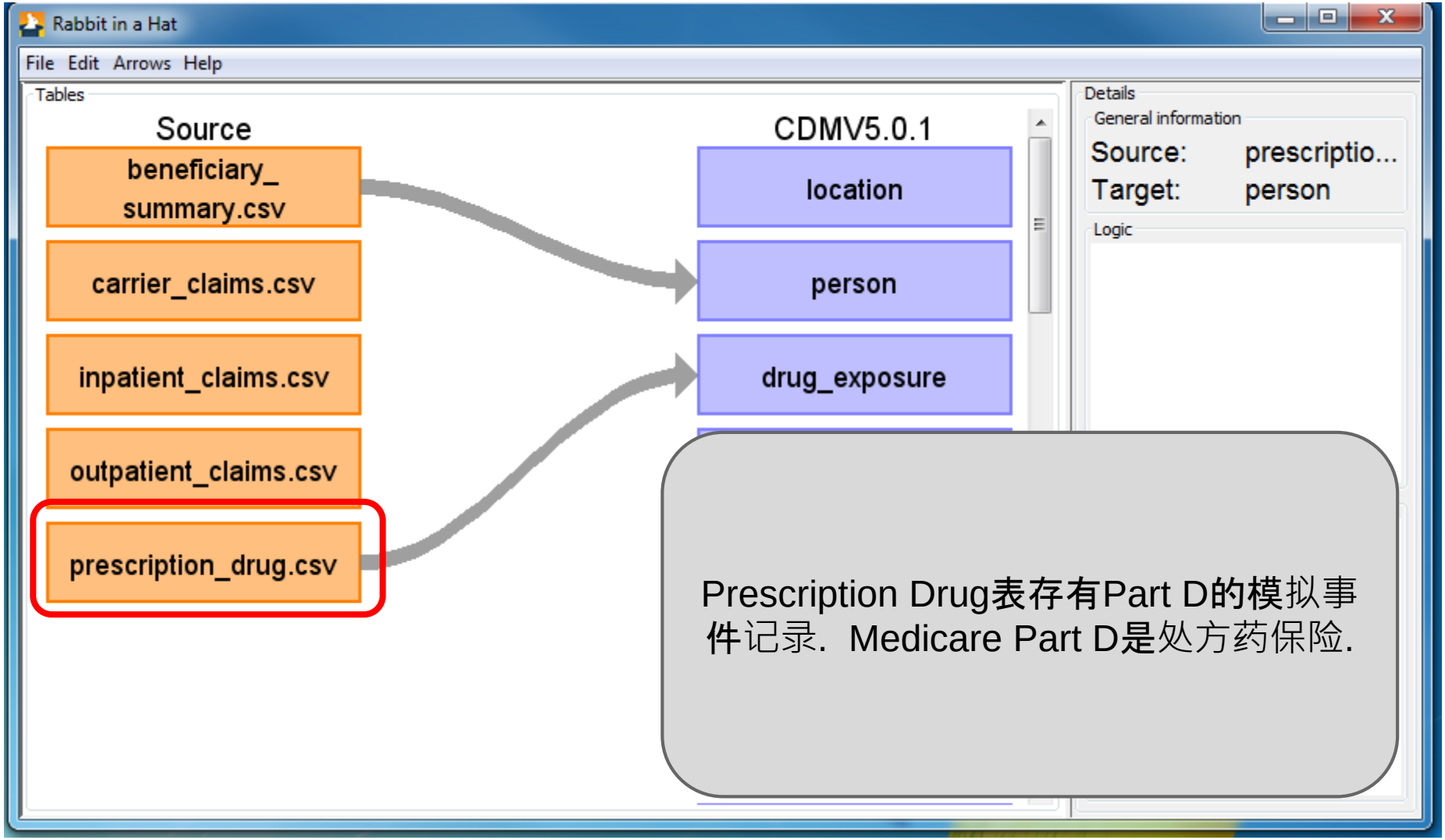
映射原始表至CDM表


Rabbit in a Hat
File Edit Arrows Help
Tables
Source
beneficiary_ summary.csv
carrier_claims.csv
inpatient_claims.csv
outpatient_claims.csv
prescription_drug.csv
CDMV5.0.1
location
person
observation_period
care_site
visit_occurrence
provider
condition_occurrence
Details

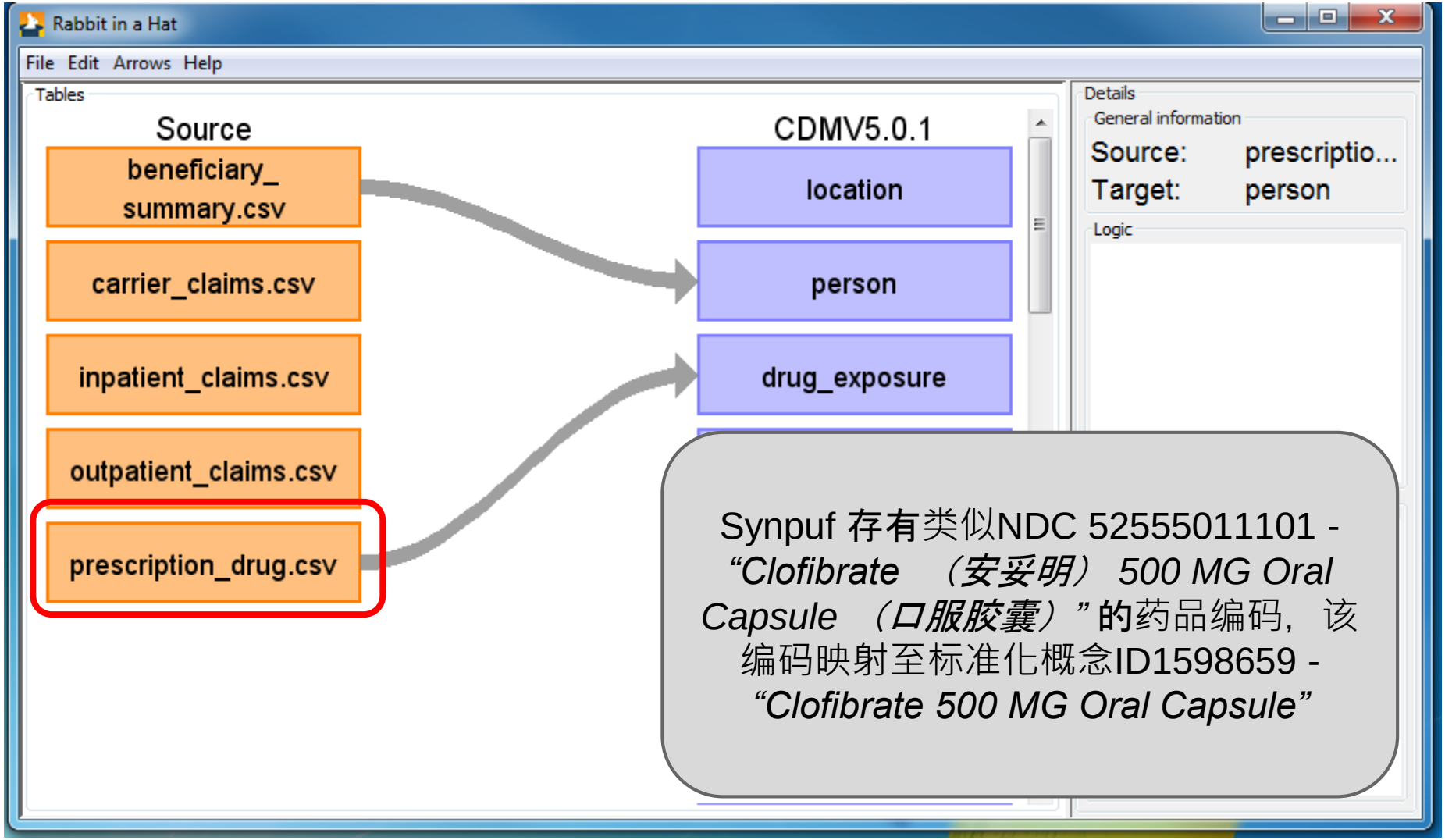
映射原始表至CDM表


Rabbit in a Hat
File Edit Arrows Help
Tables
Source
CDMV5.0.1
Details
beneficiary_ summary.csv
carrier_claims.csv
inpatient_claims.csv
outpatient_claims.csv
prescription_drug.csv
location
person
drug_exposure
observation_period
care_site
visit_occurrence
provider

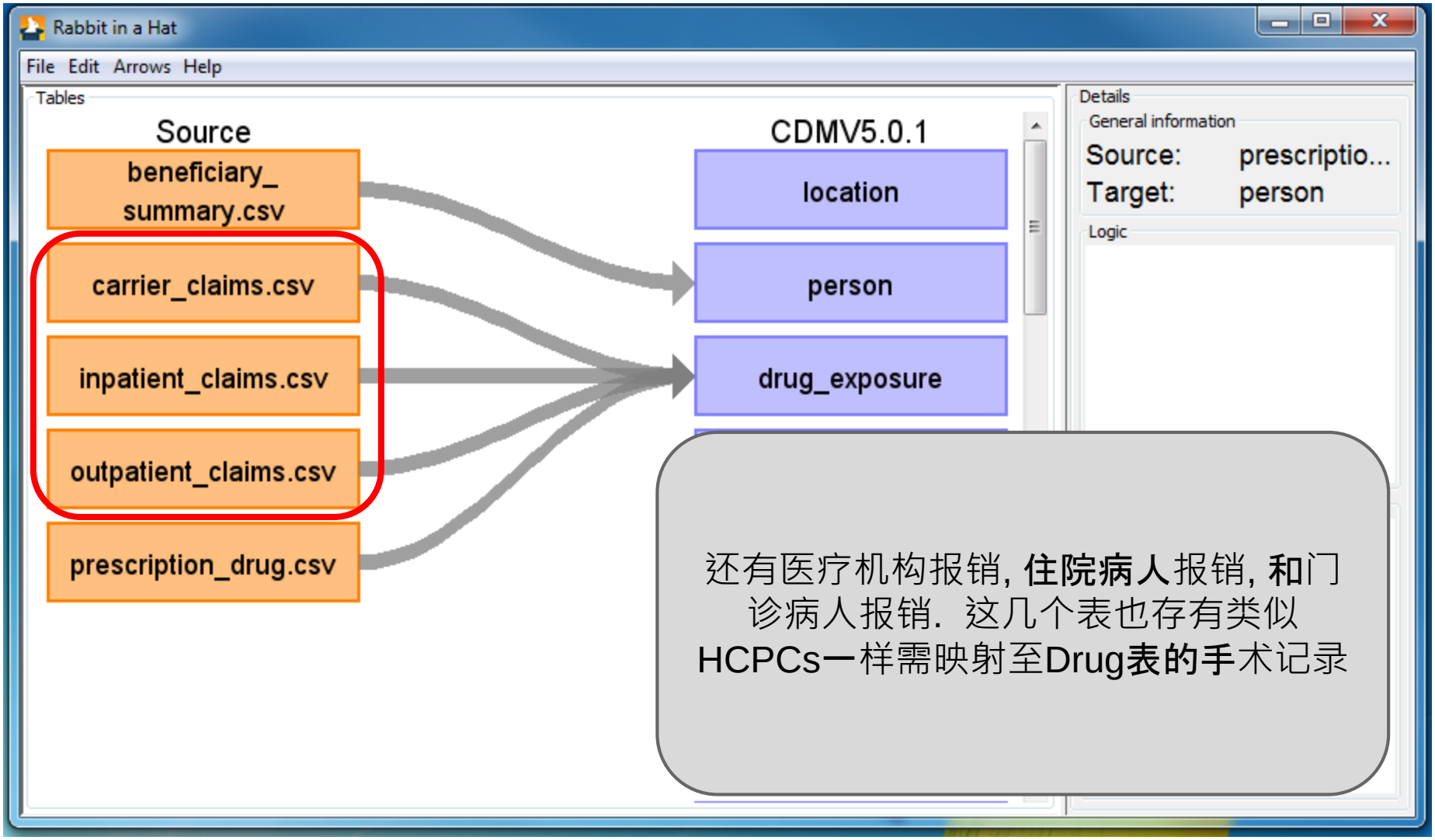
映射原始表至CDM表


Rabbit in a Hat
File Edit Arrows Help
Tables
Source
beneficiary_ summary.csv
carrier_claims.csv
inpatient_claims.csv
outpatient_claims.csv
prescription_drug.csv
CDMV5.0.1
location
person
drug_exposure
Details
General information
Source: prescriptio...
Target: person
Logic
Prescription Drug表存有Part D的模拟事件记录. Medicare Part D是处方药保险.

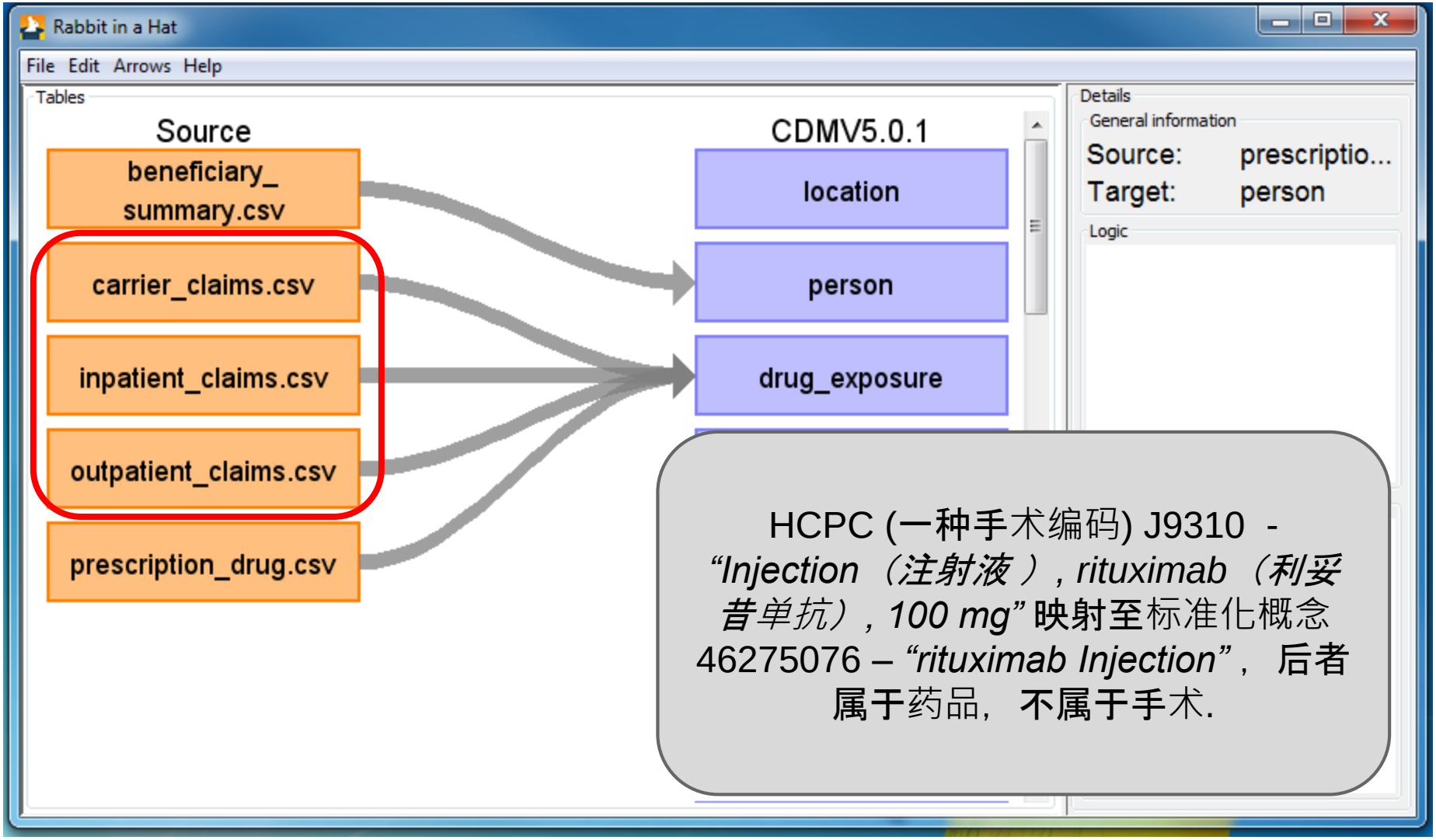
映射原始表至CDM表


Rabbit in a Hat
File Edit Arrows Help
Tables
Source
beneficiary_ summary.csv
carrier_claims.csv
inpatient_claims.csv
outpatient_claims.csv
prescription_drug.csv
CDMV5.0.1
location
person
drug_exposure
Details
General information
Source: prescriptio...
Target: person
Logic
Synpuf 存有类似NDC 52555011101 - "Clofibrate （安妥明） 500 MG Oral Capsule （口服胶囊）" 的药品编码，该编码映射至标准化概念ID1598659 - "Clofibrate 500 MG Oral Capsule"

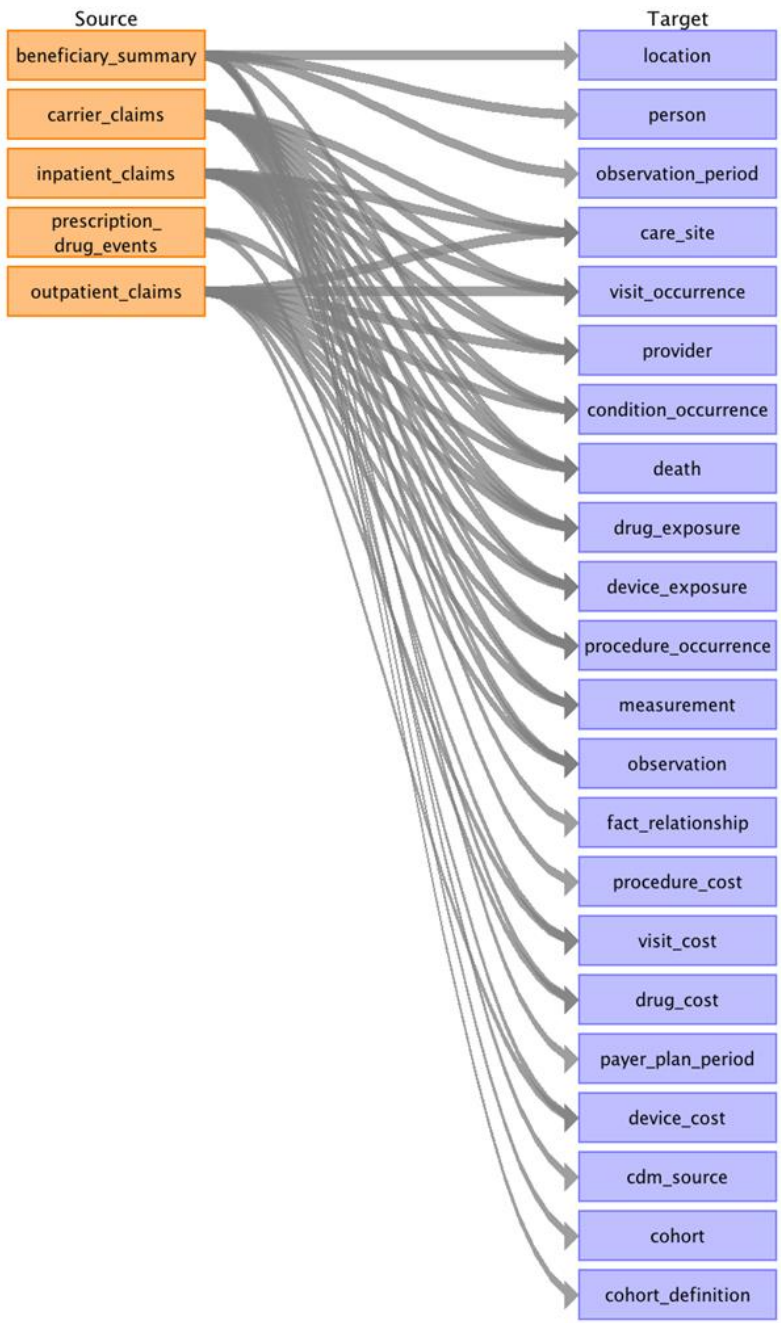
映射原始表至CDM表


Rabbit in a Hat
File Edit Arrows Help
Tables
Source
beneficiary_ summary.csv
carrier_claims.csv
inpatient_claims.csv
outpatient_claims.csv
prescription_drug.csv
CDMV5.0.1
location
person
drug_exposure
Details
General information
Source: prescriptio...
Target: person
Logic
还有医疗机构报销, 住院病人报销, 和门诊病人报销. 这几个表也存有类似HCPCs一样需映射至Drug表的手术记录

Map Raw Tables to CDM Tables



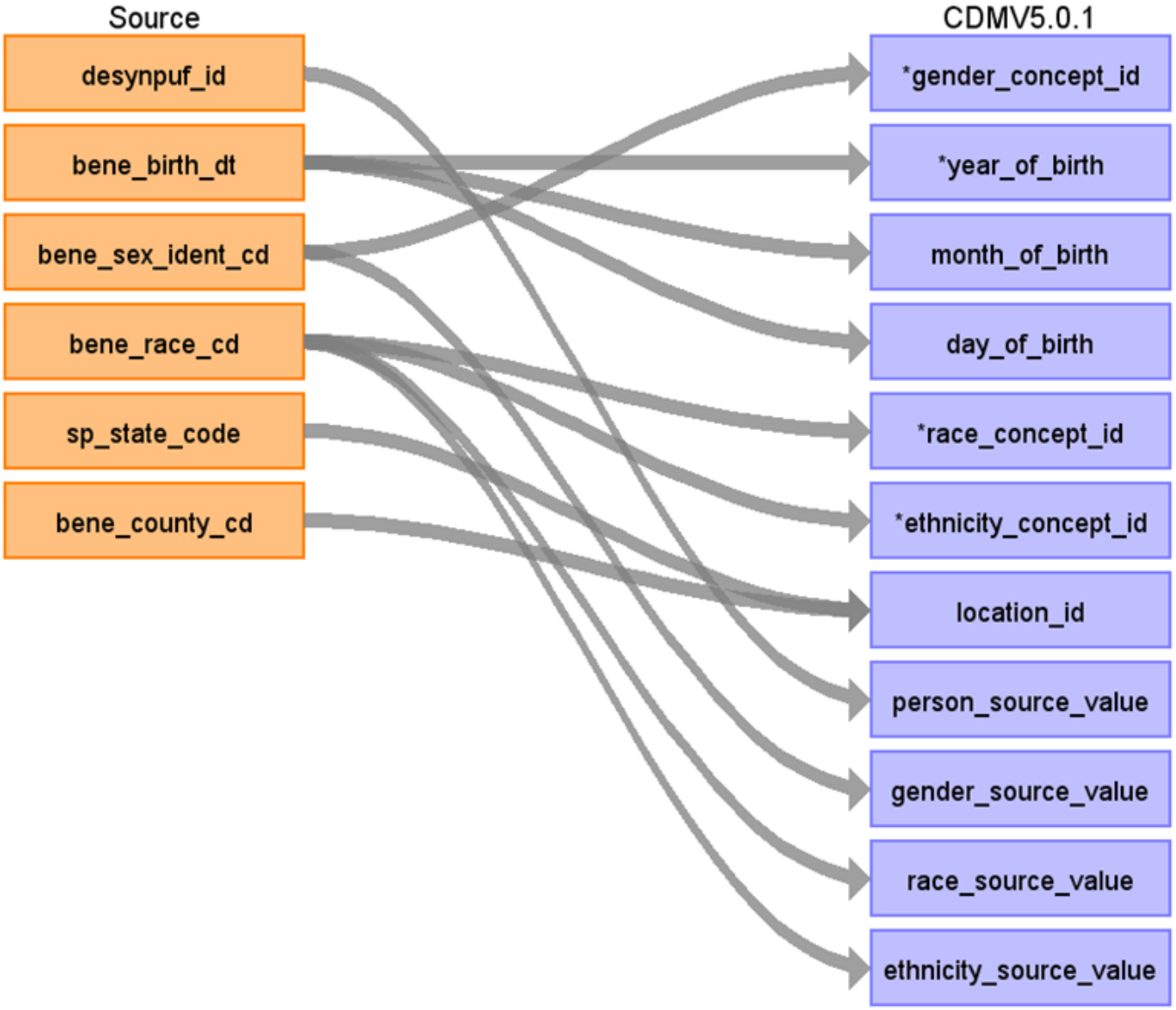
映射原始表至CDM表



继续映射原始数据表至CDM，尽量保留最多的原始数据信息

PERSON

- 我们今天的举例从PERSON 表开始


Source
desynpuf_id
bene_birth_dt
bene_sex_ident_cd
bene_race_cd
sp_state_code
bene_county_cd
CDMV5.0.1
*gender_concept_id
*year_of_birth
month_of_birth
day_of_birth
*race_concept_id
*ethnicity_concept_id
location_id
person_source_value
gender_source_value
race_source_value
ethnicity_source_value

Destination Field	Source Field	Logic	Comment
person_id			Autonumber
gender_concept_id	bene_sex_ident_cd	Source Value - Standard Concept Id 1 - 8507 2 - 8532 If gender is not 1 or 2, please discard person.	1-Male 2-Female
year_of_birth	bene_birth_dt	Take first 4 digits (starting from left)	BENE_BIRTH_DT = YYYYMMDD
month_of_birth	bene_birth_dt	Take 5th and 6th digit starting from the left	BENE_BIRTH_DT = YYYYMMDD
day_of_birth	bene_birth_dt	Take last two digits starting from the left.	BENE_BIRTH_DT = YYYYMMDD
time_of_birth			N/A
race_concept_id	bene_race_cd	Source Value - Concept ID 1 - 8527 2 - 8516 3 - 0 5 - 0 Else set to 0.	1-White 2-Black 3-Others 5-Hispanic
ethnicity_concept_id	bene_race_cd	Source Value - Concept ID 1 - 38003564 2 - 38003564 3 - 0 5 - 38003563 Else set to 0.	1-White 2-Black 3-Others 5-Hispanic
location_id	sp_state_code bene_county_cd	Use the BENE_COUNTY_CD and SP_STATE_CODE to lookup in the LOCATION table the LOCATION_ID.	
provider_id			N/A
care_site_id			N/A
person_source_value	desynpuf_id		
gender_source_value	bene_sex_ident_cd		
gender_source_concept_id			Set to 0.
race_source_value	bene_race_cd		
race_source_concept_id			Set to 0.
ethnicity_source_value	bene_race_cd		
ethnicity_source_concept_id			Set to 0.


OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

应用

使用 Usagi

Github

https://github.com/OHDSI/Usagi

要求Java 1.8 及以上

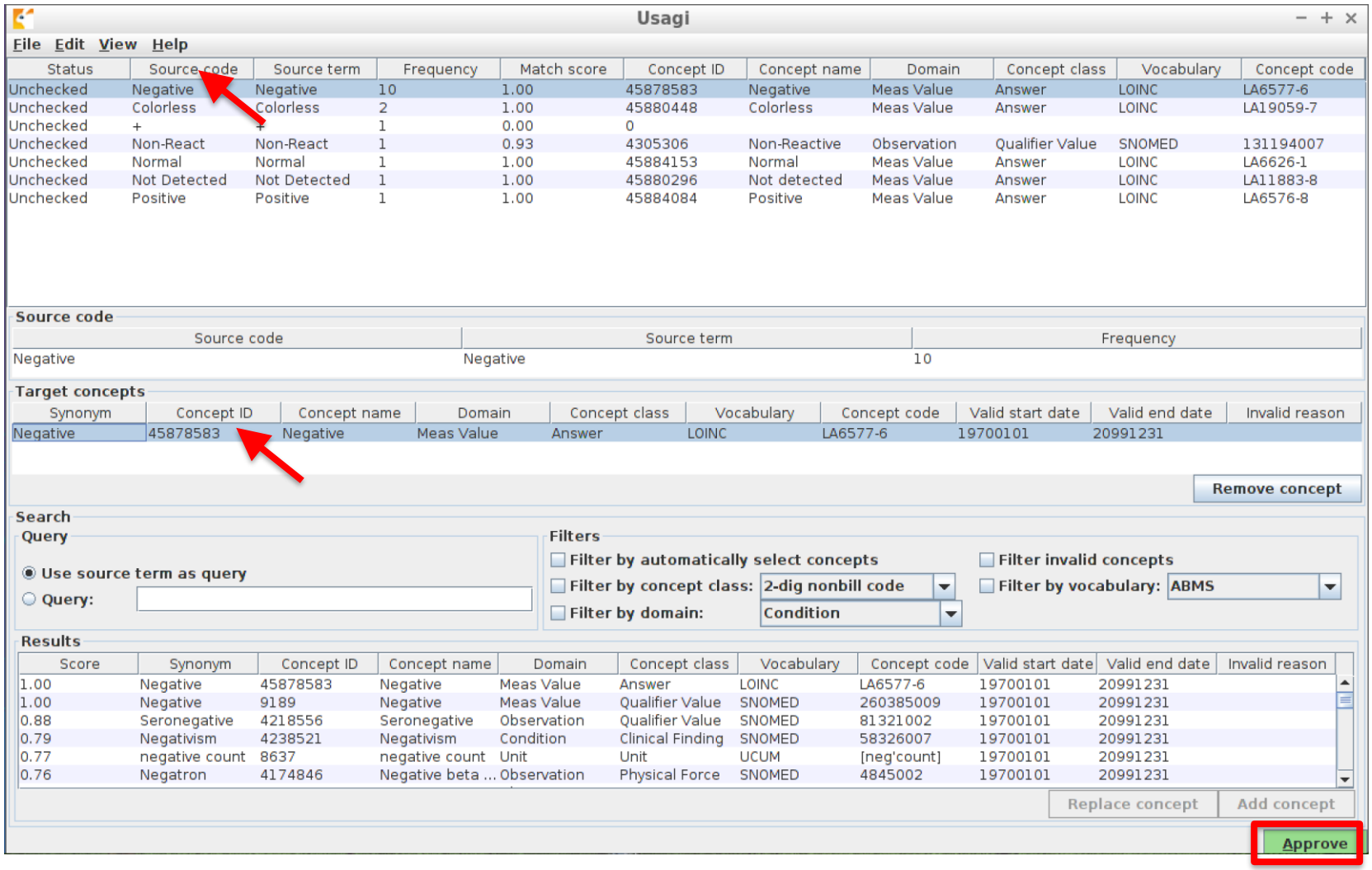
- 下载最新版 Usagi jar 文件
- 需要安装 v5 术语文件

USAGI

- 用于帮助将源系统中的代码映射到存储在OMOP词汇表中的标准术语的工具

http://www.ohdsi.org/web/wiki/doku.php?id=documentation:software:usagi

USAGI 练习


Usagi
File Edit View Help
Status | Source code | Source term | Frequency | Match score | Concept ID | Concept name | Domain | Concept class | Vocabulary | Concept code
Unchecked | Negative | Negative | 10 | 1.00 | 45878583 | Negative | Meas Value | Answer | LOINC | LA6577-6
Unchecked | Colorless | Colorless | 2 | 1.00 | 45880448 | Colorless | Meas Value | Answer | LOINC | LA19059-7
Unchecked | + | + | 1 | 0.00 | 0
Unchecked | Non-React | Non-React | 1 | 0.93 | 4305306 | Non-Reactive | Observation | Qualifier Value | SNOMED | 131194007
Unchecked | Normal | Normal | 1 | 1.00 | 45884153 | Normal | Meas Value | Answer | LOINC | LA6626-1
Unchecked | Not Detected | Not Detected | 1 | 1.00 | 45880296 | Not detected | Meas Value | Answer | LOINC | LA11883-8
Unchecked | Positive | Positive | 1 | 1.00 | 45884084 | Positive | Meas Value | Answer | LOINC | LA6576-8
Source code
Source code | Source term | Frequency
Negative | Negative | 10
Target concepts
Synonym | Concept ID | Concept name | Domain | Concept class | Vocabulary | Concept code | Valid start date | Valid end date | Invalid reason
Negative | 45878583 | Negative | Meas Value | Answer | LOINC | LA6577-6 | 19700101 | 20991231
Remove concept
Search
Query
Use source term as query
Query:
Filters
Filter by automatically select concepts
Filter by concept class: 2-dig nonbill code
Filter by domain: Condition
Filter invalid concepts
Filter by vocabulary: ABMS
Results
Score | Synonym | Concept ID | Concept name | Domain | Concept class | Vocabulary | Concept code | Valid start date | Valid end date | Invalid reason
1.00 | Negative | 45878583 | Negative | Meas Value | Answer | LOINC | LA6577-6 | 19700101 | 20991231
1.00 | Negative | 9189 | Negative | Meas Value | Qualifier Value | SNOMED | 260385009 | 19700101 | 20991231
0.88 | Seronegative | 4218556 | Seronegative | Observation | Qualifier Value | SNOMED | 81321002 | 19700101 | 20991231
0.79 | Negativism | 4238521 | Negativism | Condition | Clinical Finding | SNOMED | 58326007 | 19700101 | 20991231
0.77 | negative count | 8637 | negative count | Unit | Unit | UCUM | [neg'count] | 19700101 | 20991231
0.76 | Negatron | 4174846 | Negative beta ... | Observation | Physical Force | SNOMED | 4845002 | 19700101 | 20991231
Replace concept
Add concept
Approve

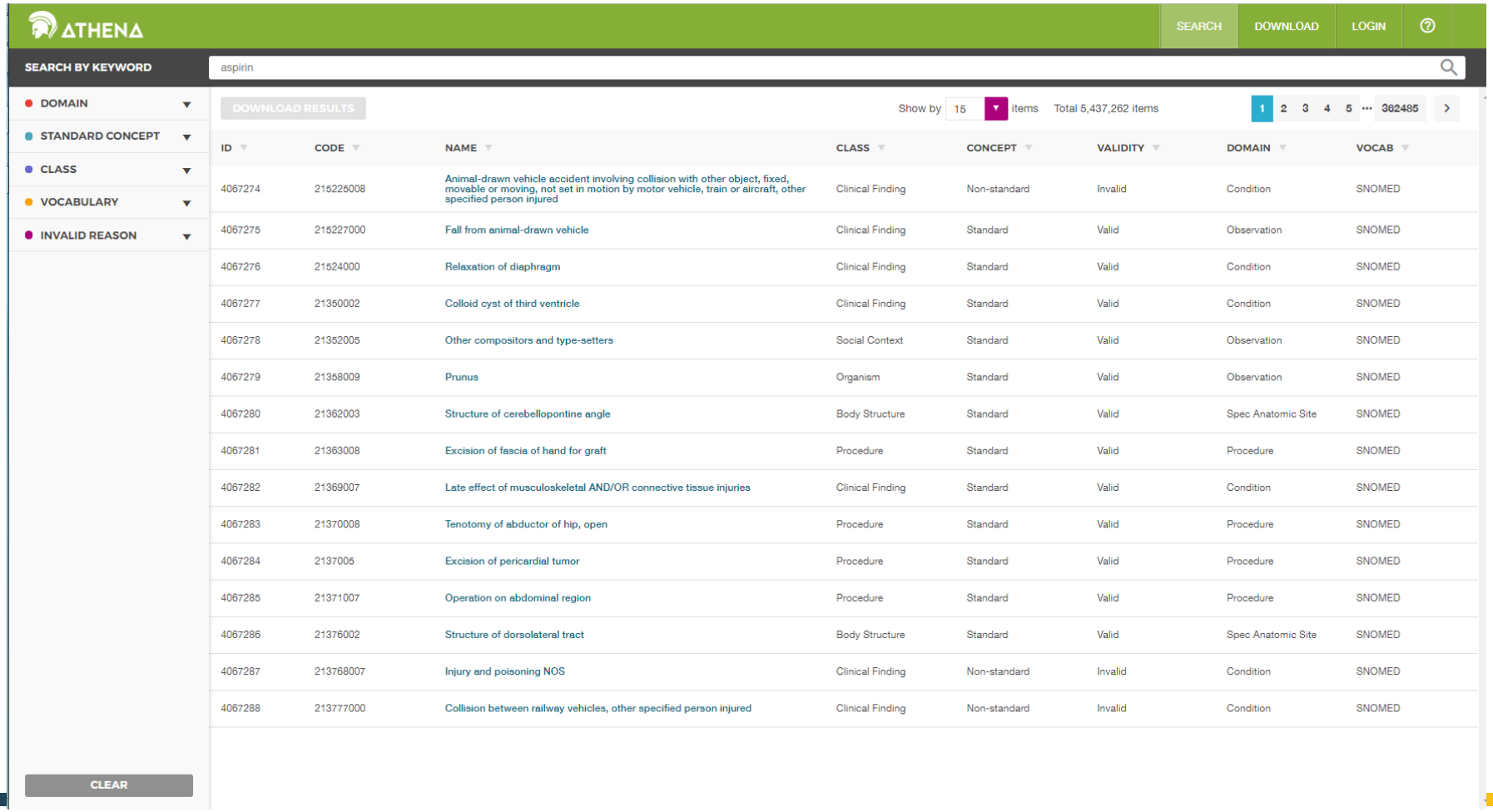

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

应用

使用 Athena

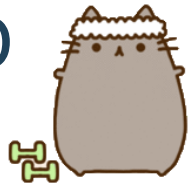
Github

http://athena.ohdsi.org/

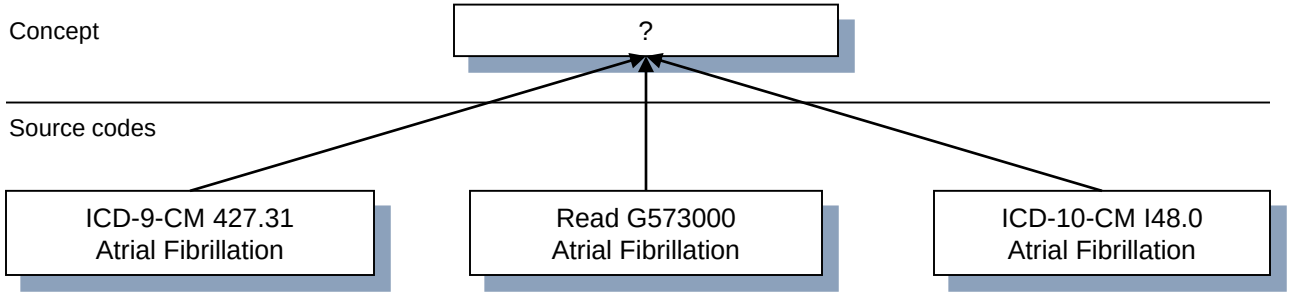


ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
4067274	215225008	Animal-drawn vehicle accident involving collision with other object, fixed, movable or moving, not set in motion by motor vehicle, train or aircraft, other specified person injured	Clinical Finding	Non-standard	Invalid	Condition	SNOMED
4067275	215227000	Fall from animal-drawn vehicle	Clinical Finding	Standard	Valid	Observation	SNOMED
4067276	21524000	Relaxation of diaphragm	Clinical Finding	Standard	Valid	Condition	SNOMED
4067277	21350002	Colloid cyst of third ventricle	Clinical Finding	Standard	Valid	Condition	SNOMED
4067278	21352005	Other compositors and type-setters	Social Context	Standard	Valid	Observation	SNOMED
4067279	21358009	Prunus	Organism	Standard	Valid	Observation	SNOMED
4067280	21362003	Structure of cerebellopontine angle	Body Structure	Standard	Valid	Spec Anatomic Site	SNOMED
4067281	21363008	Excision of fascia of hand for graft	Procedure	Standard	Valid	Procedure	SNOMED
4067282	21369007	Late effect of musculoskeletal AND/OR connective tissue injuries	Clinical Finding	Standard	Valid	Condition	SNOMED
4067283	21370008	Tenotomy of abductor of hip, open	Procedure	Standard	Valid	Procedure	SNOMED
4067284	2137005	Excision of pericardial tumor	Procedure	Standard	Valid	Procedure	SNOMED
4067285	21371007	Operation on abdominal region	Procedure	Standard	Valid	Procedure	SNOMED
4067286	21376002	Structure of dorsolateral tract	Body Structure	Standard	Valid	Spec Anatomic Site	SNOMED
4067287	213768007	Injury and poisoning NOS	Clinical Finding	Non-standard	Invalid	Condition	SNOMED
4067288	213777000	Collision between railway vehicles, other specified person injured	Clinical Finding	Non-standard	Invalid	Condition	SNOMED

Exercise: Find Standard Concept ID from Source Concept



```
ICD-9: '427.31'    : 313217
Read: 'G573000'    : 313217
ICD-10: 'I48.0'    : 4154290 'Paroxysmal Atrial Fibrillation'
```



Step 1. Lookup

```
SELECT * FROM concept WHERE concept_code = ...;
```

Step 2. Translate

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

Step 3. Check out

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

Exercise: Find Standard Concept ID for Conditions

- Asthma — 317009
- Plague — 434271
- Ingrown toenail — 4065236, 4290993
- Your favorite condition here

Exercise:
Find Standard Concept ID

- Metformin 1503297
- Tolazamide 1502809
- Telmisartan 1317640
- Your favorite ingredient here

Exercise:
Find Standard Concept ID

- A10AE06 — 35602717
- 686450400 — 19080217
- Your favorite drug here

Exercise:
Log into Atlas

URL:
https://training.atlasplus.imshealth.com/#/home

Username: Atlasuser2@gmail.com

Password: Trainer@2018

OHDSI 教程：设计和实现

Patrick Ryan, PhD

Janssen Research and Development, Columbia University Medical Center

Jon Duke, MD

Georgia Tech Research Institute

Agenda

Section	Time	Item(s)
Cohort Definition	8:00 – 9:00 (1 hour)	Index date Entry criteria Additional criteria Exit criteria
Atlas Demo	9:00 – 10:00 (1 hour)	Data Source Profile Vocabulary Concept Set
Break	10:00 – 10:15	Break
Cohort Building	10:15 – 12:00 (45 min)	Graham paper Cohort Creation
Lunch	12:00 – 13:00	Lunch
Cohort Building	13:00 – 14:00	Cohort Features

Agenda (cont)

Section	Time	Item(s)
Population Estimation	14:00 – 14:30 (30 minutes)	Target, Comparator, Outcome Cohort
Break	14:30 – 14:45	Break
Population Estimation	14:45 – 17:00 (45 min)	Build Population Estimation Review Results

OHDSI's mission

To improve health, by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

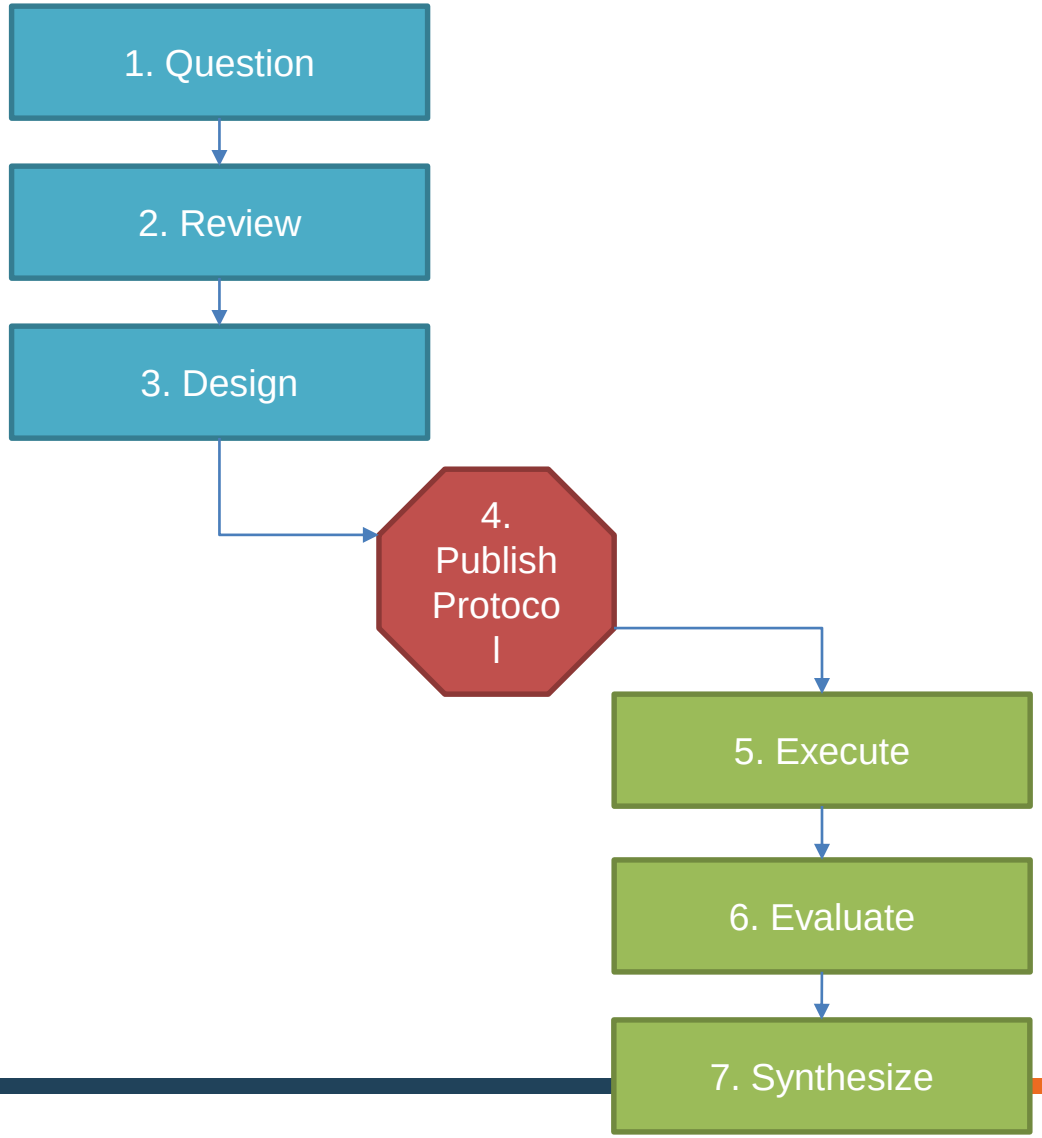
What evidence does OHDSI seek to generate from observational data?

- Clinical characterization
  - **Natural history:** Who are the patients who have diabetes? Among those patients, who takes metformin?
  - **Quality improvement:** what proportion of patients with diabetes experience disease-related complications?
- Population-level estimation
  - **Safety surveillance:** Does metformin cause lactic acidosis?
  - **Comparative effectiveness:** Does metformin cause lactic acidosis more than glyburide?
- Patient-level prediction
  - **Precision medicine:** Given everything you know about me and my medical history, if I start taking metformin, what is the chance that I am going to have lactic acidosis in the next year?
  - **Disease interception:** Given everything you know about me, what is the chance I will develop diabetes?

What is OHDSI's strategy to deliver reliable evidence?

- **Methodological research**
  - Develop new approaches to observational data analysis
  - Evaluate the performance of new and existing methods
  - Establish empirically-based scientific best practices
- **Open-source analytics development**
  - Design tools for data transformation and standardization
  - Implement statistical methods for large-scale analytics
  - Build interactive visualization for evidence exploration
- **Clinical evidence generation**
  - Identify clinically-relevant questions that require real-world evidence
  - Execute research studies by applying scientific best practices through open-source tools across the OHDSI international data network
  - Promote open-science strategies for transparent study design and evidence dissemination

A standardized process for evidence generation and dissemination



How OHDSI is trying to help:

OHDSI community

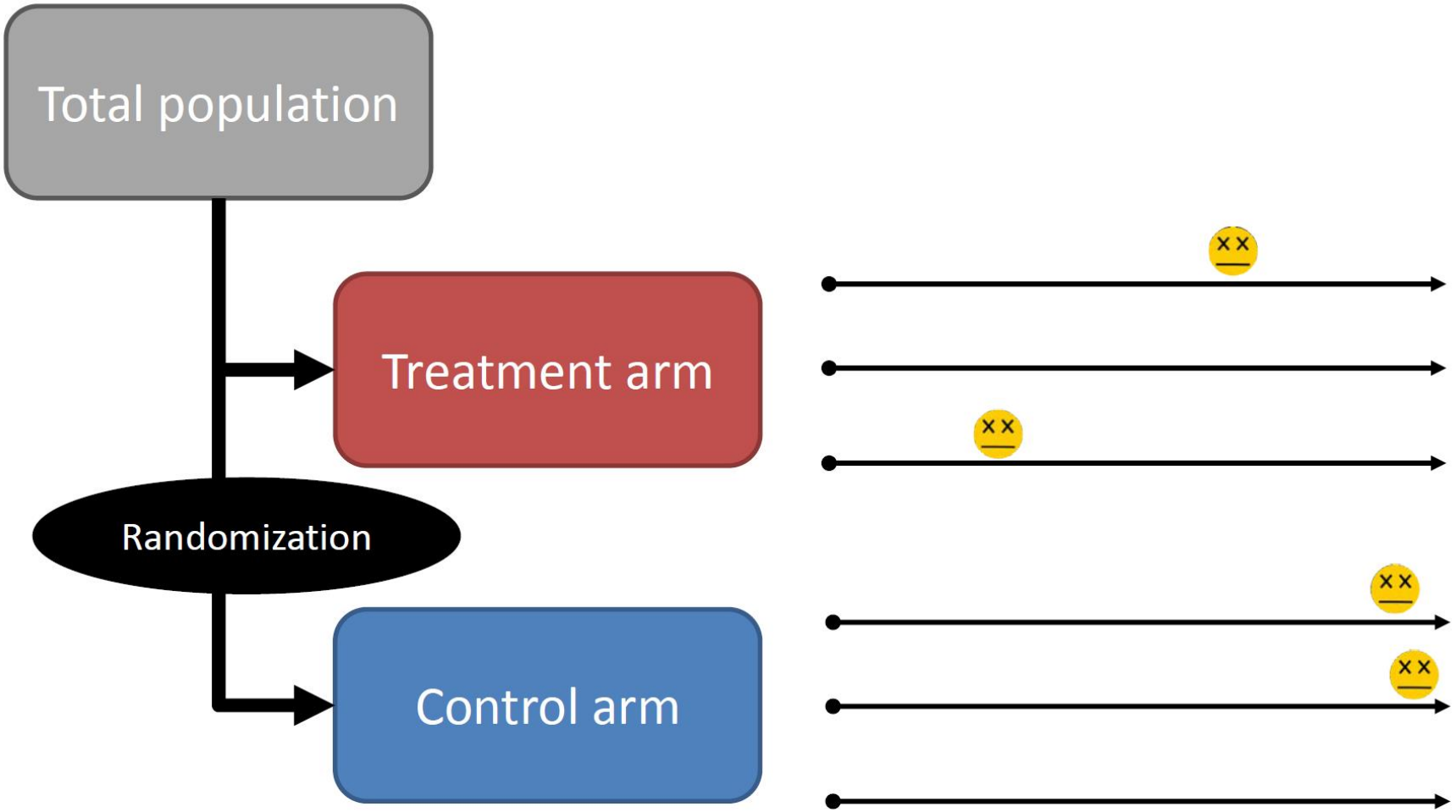
Open-source knowledgebase (LAERTES)

Open-source front-end web applications (ATLAS)

Open-source back-end statistical packages
(R Methods Library)

OHDSI network studies

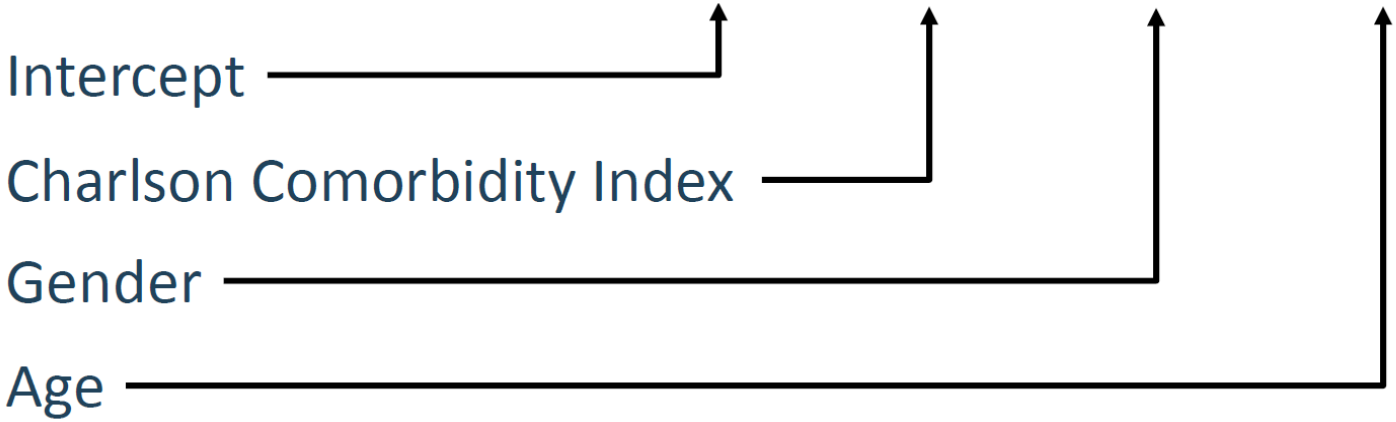
RCT VS Comparative Cohort Study



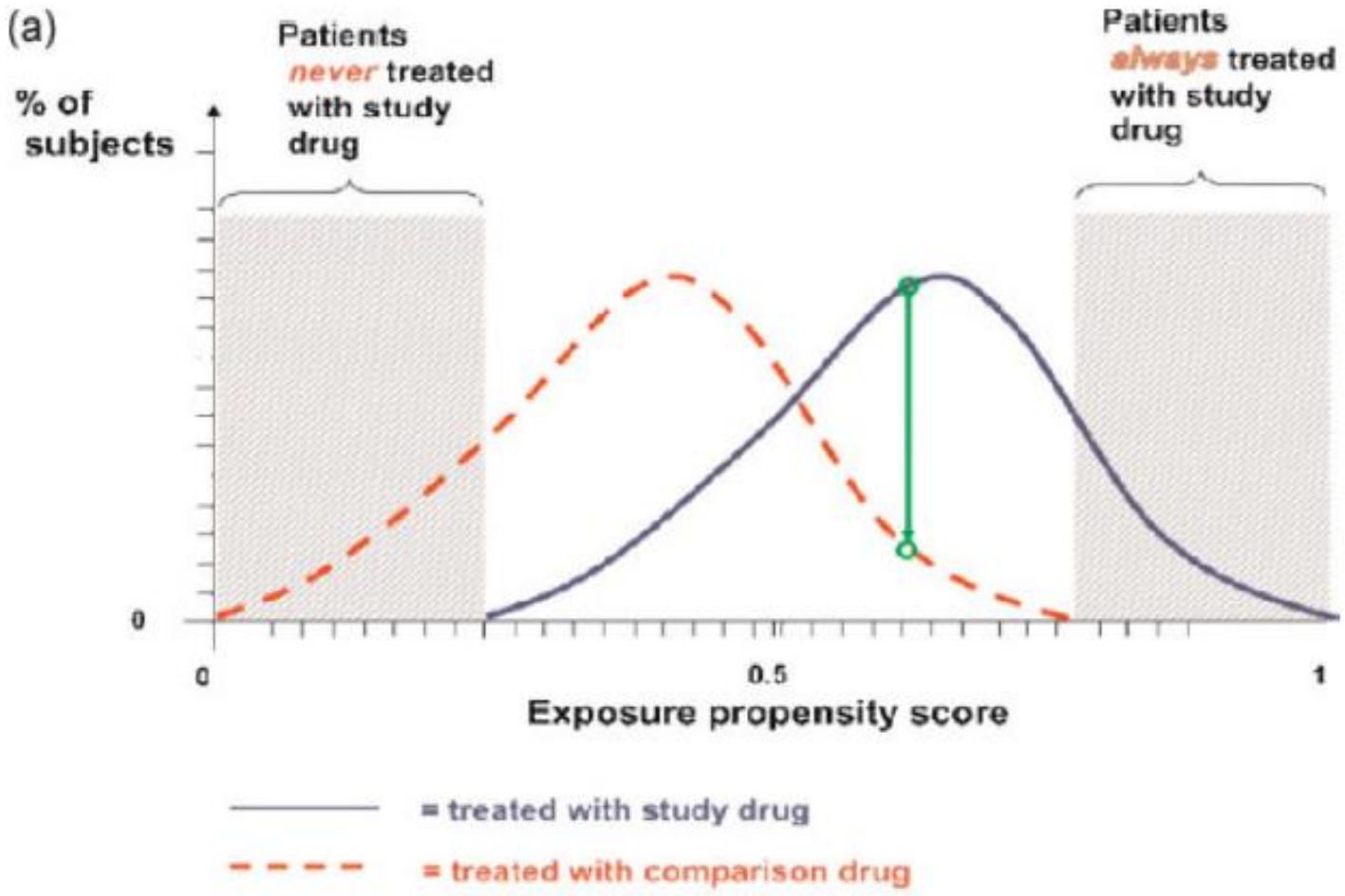
Propensity score (PS)

The propensity score is the probability of receiving the treatment, conditional on a set of baseline characteristics

$$P(\,treatment \mid X) = f(\,\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots\,)$$



Intuition around propensity score balance



Schneeweiss. PDS 2011

Methods for confounding adjustment using a propensity score

Regression adjustment	The PS is used as a covariable in an outcome regression model to adjust the as[illegible] assum[illegible]e same [illegible] relationship between propensity score and outcome is correctly specified.
Matching	The PS is used to match exposed subjects to unexposed subjects with similar values of the PS. This method assumes that within the matched sample, exposed and unexposed subjects have a similar distribution of baseline characteristics.
Stratification	The PS is used to stratify subjects into (often quintiles or deciles) strata. Treatment effects are estimated separately within each stratum and then combined into an overall estimate of treatment effect. This method assumes that within each stratum, exposed and unexposed subjects have a similar distribution of baseline characteristics.
Inverse Probability Weighting	The PS is used to create weights based on the inverse probability which is defined as: E*/PS + (1-E)/(1-PS). This assumes that baseline characteristics are similar in the exposed and unexposed group.

* E: exposure

Not generally recommended

Fully implemented in OHDSI CohortMethod R package

Garbe et al, Eur J Clin Pharmacol 2013, http://www.ncbi.nlm.nih.gov/pubmed/22763756

How does Epidemiology define a comparative cohort study?

…it depends on what Epidemiology textbook you read…

"In a retrospective cohort study…the investigator identified the cohort of individu
their su
recent

"Cohort studies are studies that identify subsets of a defined population and follow them over time, looking for differences in their outcome. Cohort studies generally compare exposed patients to unexposed patients, although they can also be used to compare one exposure to another."
--Strom, Pharmacoepidemiology, 2005

"In
has experienced the outcome of interest, but all of whom could experience
it…On
charac
people

"In the coho
identified. I
incidence of
ascertained
--S

"In the paradigmatic cohort study, the investigator defines two or more groups of people that are free of disease and that differ according to the extent of their exposure to a potential cause of disease. These groups are referred to as the study cohorts. When two groups are studies, one is usually though of as the exposed or index cohort – those individuals who have experienced the putative causal event or condition – and the other is then thought of as the unexposed or reference cohort."
--Rothman, Modern Epidemiology, 2008

OHDSI's definition of 'cohort'
OHDSI的“队列”定义

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

队列 = 在一段时期内满足一条或多条入组标准的个体集合

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

基于该定义的目标结局:

- 同一个个体可以属于多个队列
- 在不同时间段内同一个个体可以被多次纳入同一个队列
- 在同一个时间段内同一个个体不可以被多次纳入同一个队列
- 一个队列可以有0个或多个成员
- 代码集不等于队列…

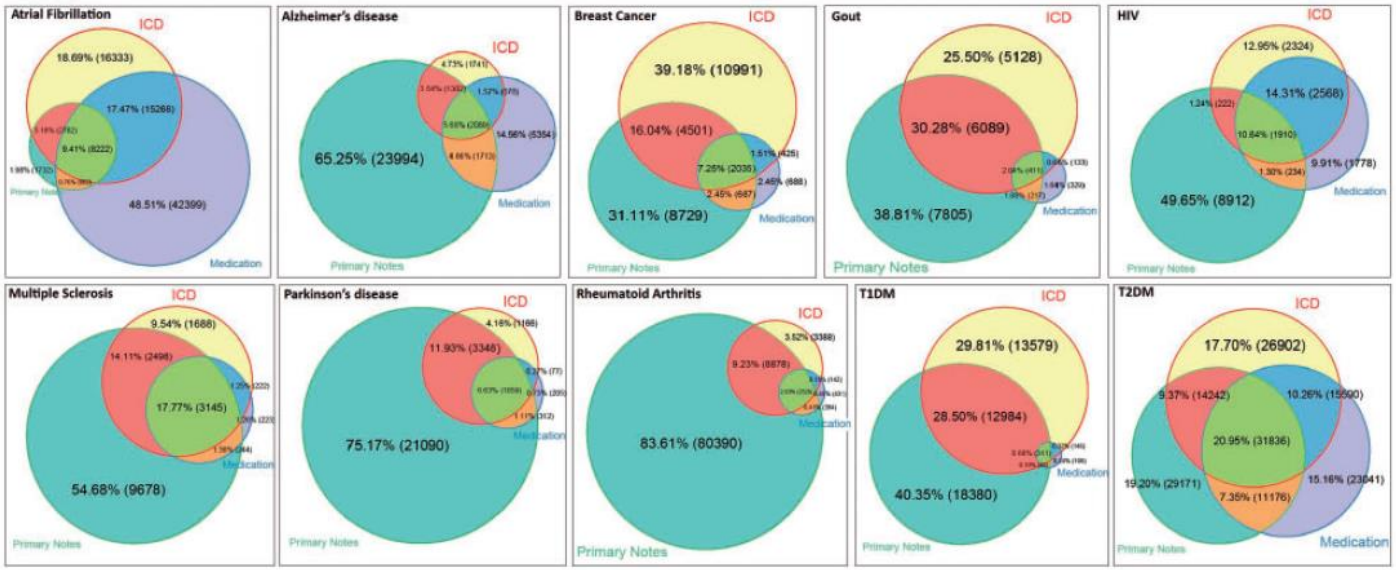
  …需要定义在标准中如何使用代码集

Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance

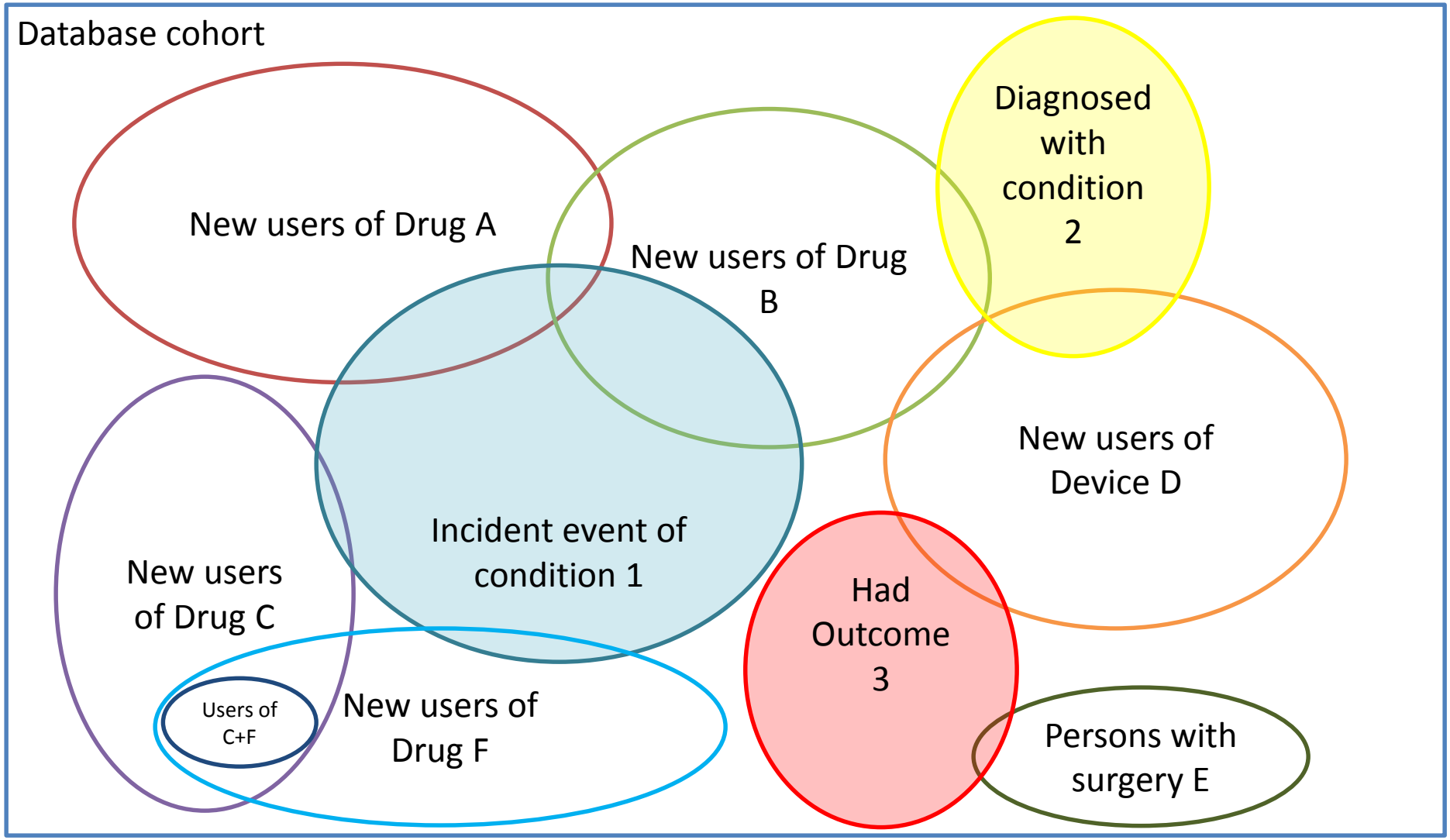
RECEIVED 8 January 2015
REVISED 14 July 2015
ACCEPTED 15 July 2015
PUBLISHED ONLINE FIRST 2 September 2015

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.



A database is full of cohorts, some of which may represent valid comparisons



An Example of Comparative Cohort Studies

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

Design of Comparative Cohort Study

Input parameter	Design choice
Target cohort (T)	dabigatran new users with prior atrial fibrillation
Comparator cohort (C)	warfarin new users with prior atrial fibrillation
Outcome cohort (O)	Ischemic stroke
Time-at-risk	1 day after cohort start → cohort end
Model specification	1:1 propensity score-matched univariable conditional Cox proportional hazards

The choice of the outcome model defines your research question

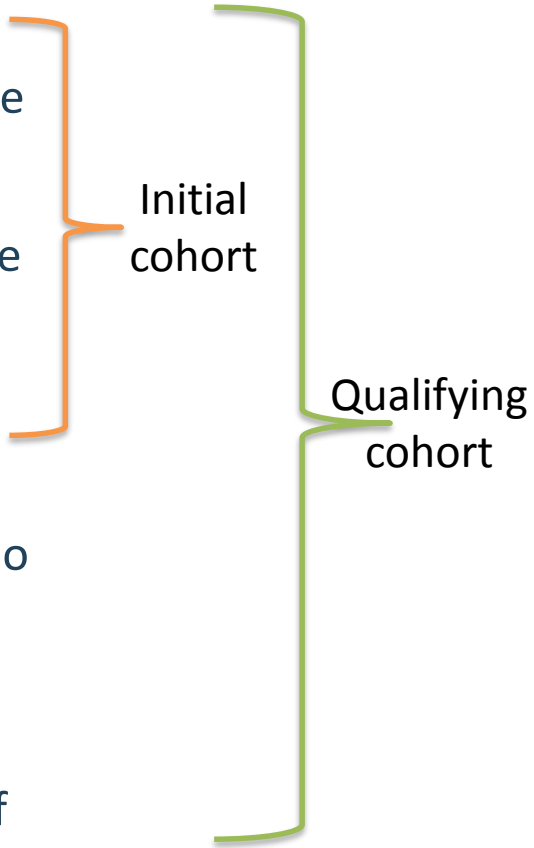
	Logistic regression	Poisson regression	Cox proportional hazards
How the outcome cohort is used	Binary classifier of presence/ absence of outcome during the fixed time-at-risk period	Count the number of occurrences of outcomes during time-at-risk	Compute time-to-event from time-at-risk start until earliest of first occurrence of outcome or time-at-risk end, and track the censoring event (outcome or no outcome)
'Risk' metric	Odds ratio	Rate ratio	Hazard ratio
Key model assumptions	Constant probability in fixed window	Outcomes follow Poisson distribution with constant risk	Proportionality – constant relative hazard


OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

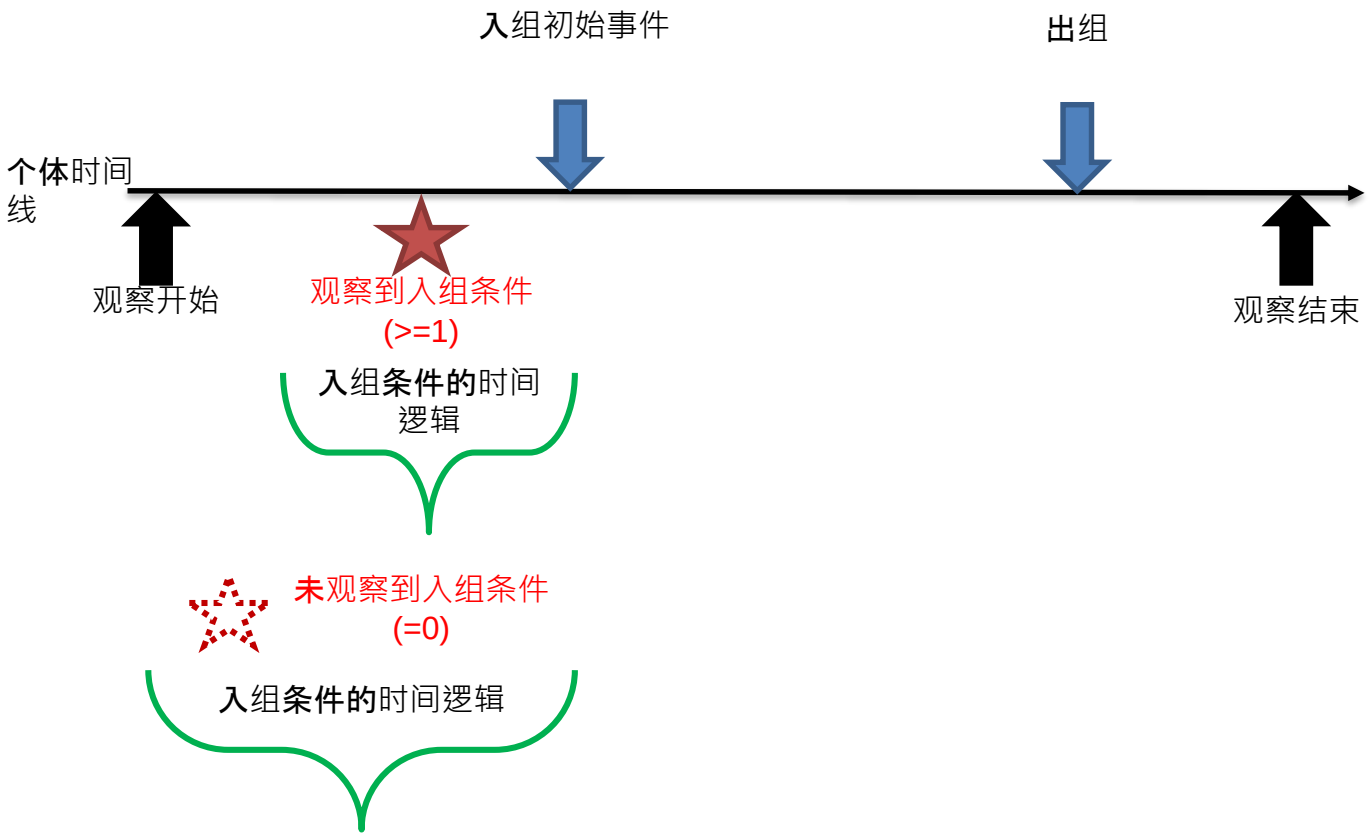
建立队列

Process flow for formally defining a cohort in ATLAS

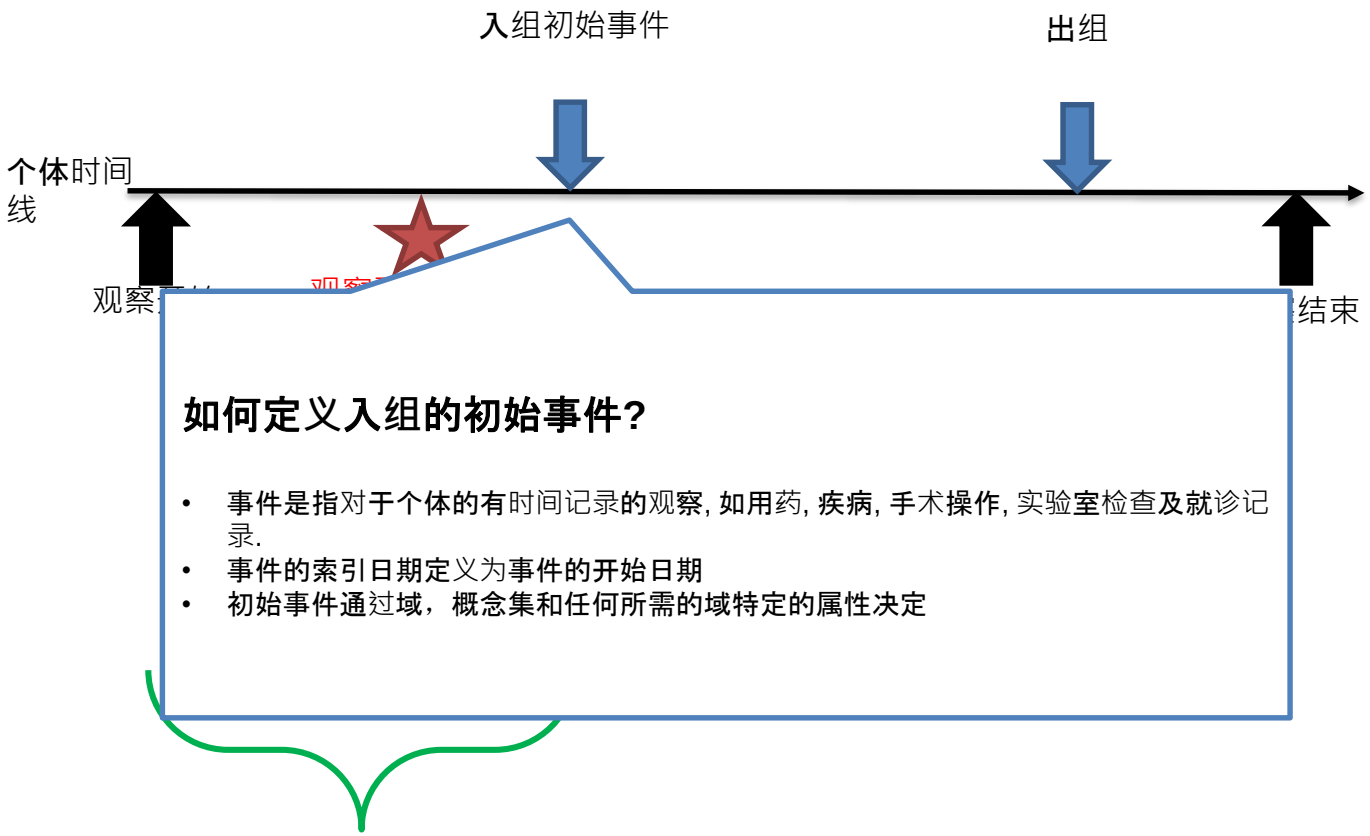
- Cohort entry criteria
  - Initial events
    - Events are recorded time-stamped observations for the persons, such as drug exposures, conditions, procedures, measurements and visits.
    - All events have a start date and end date, though some events may have a start date and end date with the same value (such as procedures or measurements).
  - Initial event inclusion criteria
  - 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.
- Cohort exit criteria



队列定义的解剖结构


入组初始事件
出组
个体时间线
观察开始
观察到入组条件
(>=1)
观察结束
入组条件的时间逻辑
未观察到入组条件
(=0)
入组条件的时间逻辑

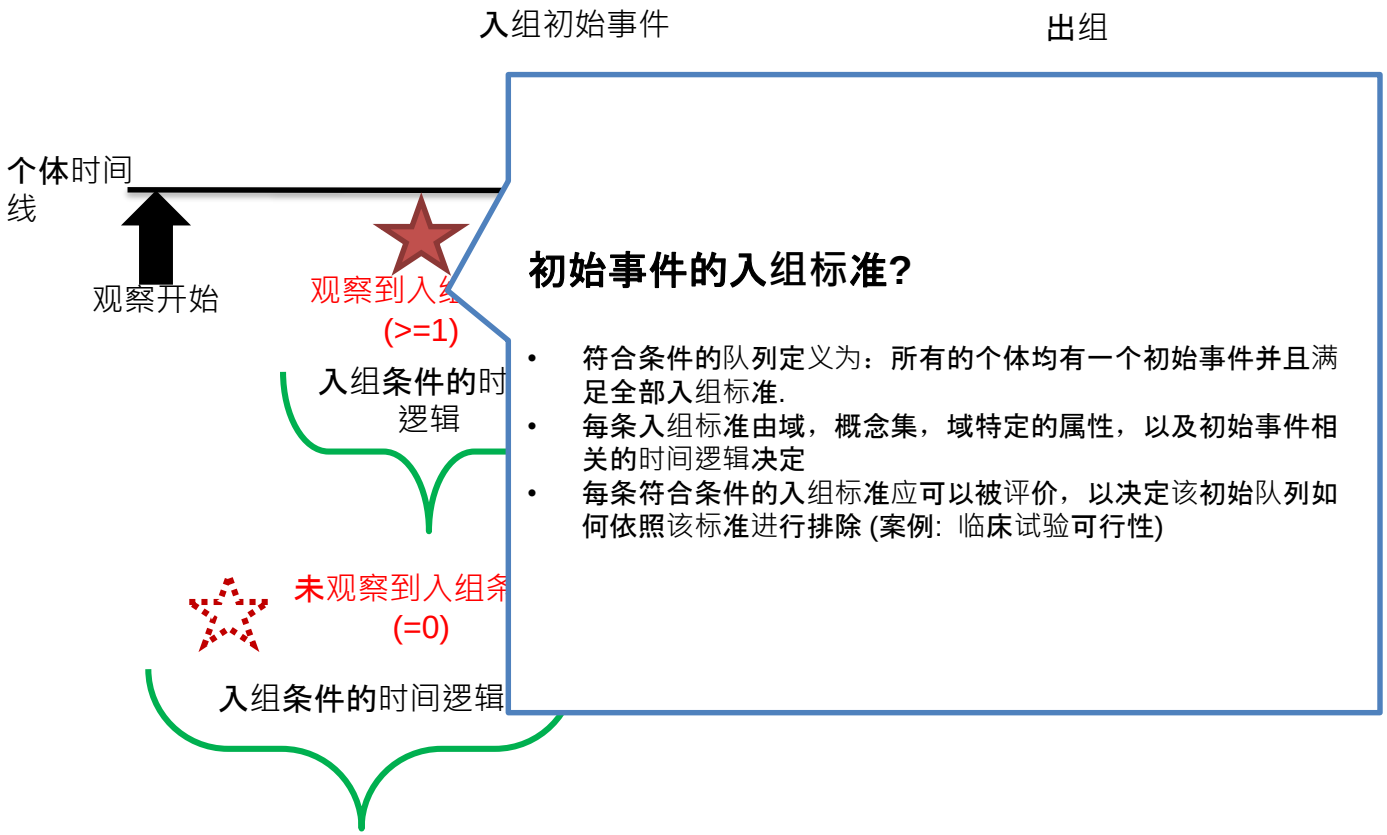
队列定义的解剖结构



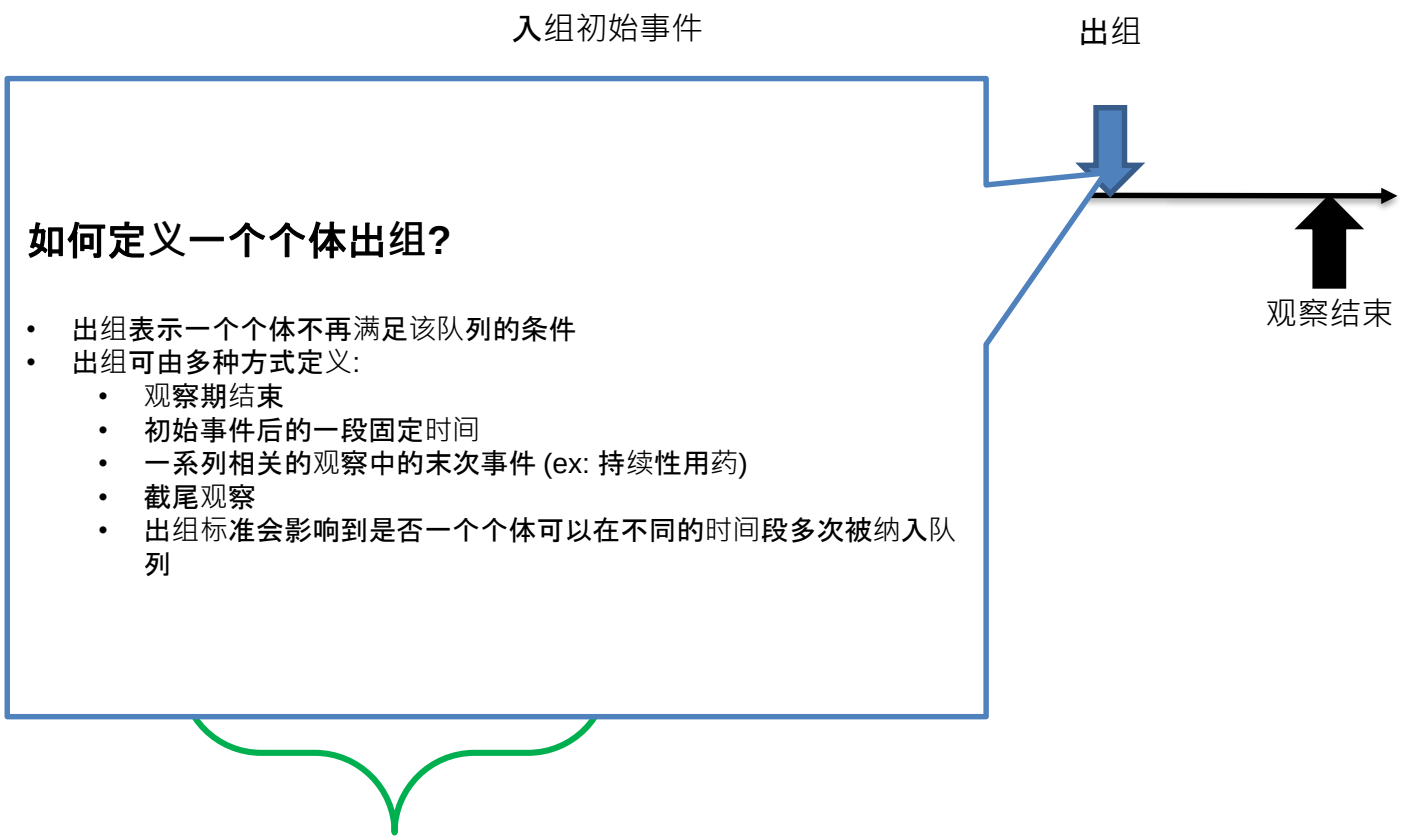
如何定义入组的初始事件?

- 事件是指对于个体的有时间记录的观察, 如用药, 疾病, 手术操作, 实验室检查及就诊记录.
- 事件的索引日期定义为事件的开始日期
- 初始事件通过域，概念集和任何所需的域特定的属性决定

队列定义的解剖结构


入组初始事件
出组
个体时间线
观察开始
观察到入组
(>=1)
入组条件的时间逻辑
未观察到入组条
(=0)
入组条件的时间逻辑
初始事件的入组标准?
• 符合条件的队列定义为：所有的个体均有一个初始事件并且满足全部入组标准.
• 每条入组标准由域，概念集，域特定的属性，以及初始事件相关的时间逻辑决定
• 每条符合条件的入组标准应可以被评价，以决定该初始队列如何依照该标准进行排除 (案例: 临床试验可行性)

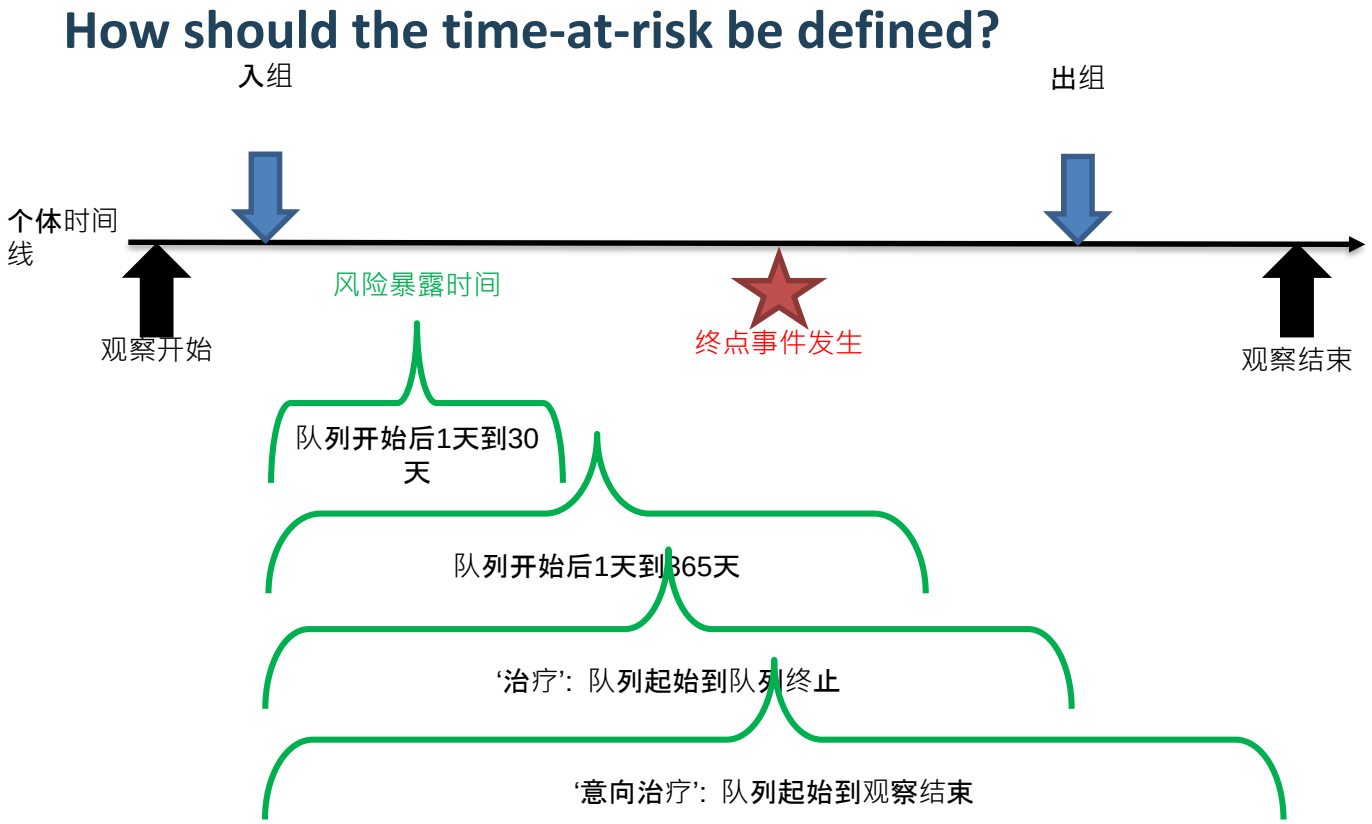
队列定义的解剖结构



如何定义一个个体出组?

- 出组表示一个个体不再满足该队列的条件
- 出组可由多种方式定义:
  - 观察期结束
  - 初始事件后的一段固定时间
  - 一系列相关的观察中的末次事件 (ex: 持续性用药)
  - 截尾观察
  - 出组标准会影响到是否一个个体可以在不同的时间段多次被纳入队列

风险暴露时间 vs 队列起始和队列终止



如何定义入组的初始事件?

- 可以:
  - 通过任何域内的阳性观察值来定义
- 不可以:
  - 通过阴性观察值来定义 – 如何确定未发生事件的时间?
  - 通过年龄定义- 出生年份是一个常数, 但是需要通过索引时间来确定年龄
- 注意:
  - 通过某一具体日期来定义队列会导致观察偏倚, 因为该日期不一定等同于接受医疗服务的日期, 例如：病例与对照匹配。可以考虑通过就诊时间段来定义。

初始事件的入组标准?

- 可以:
  - 将所有标准都列为入组标准，以避免入组标准和排除标准之间的逻辑混淆
  - 使用索引事件之前或者当天的信息
    (像随机试验一样思考: 索引事件是研究的开始，不能预测未来)
- 不可以:
  - 假设时间逻辑，却使用相关的时间窗来评价标准
- 注意:
  - “>365天观察期后首次发生”与“过去365天内无事件”是不同的
  - 一个个体可以拥有多个初始事件，入组标准应用于每个事件（而不是个体）

如何定义一个个体出组?

- 可以:
  - 明确定义出组，即使在后续分析中不会使用该定义
- 不可以:
  - 将分析中的截尾与队列的定义（该定义可以独立于分析）混淆...ex: 在终点事件发生时截尾
- 注意:
  - 参与队列的时间可以与分析时使用的风险暴露时间不同...ex: 暴露的急性影响可以采用暴露后固定时间窗进行研究, 意向治疗分析可以一直随访至观察结束

Atlas

http://ohdsi.org/web/ATLAS

讲座结束后，希望您能够

1. **熟悉ATLAS 的适用和功能**

2. **使用ATLAS** 选择有特定临床特征的病人组**并**对他们进行分析

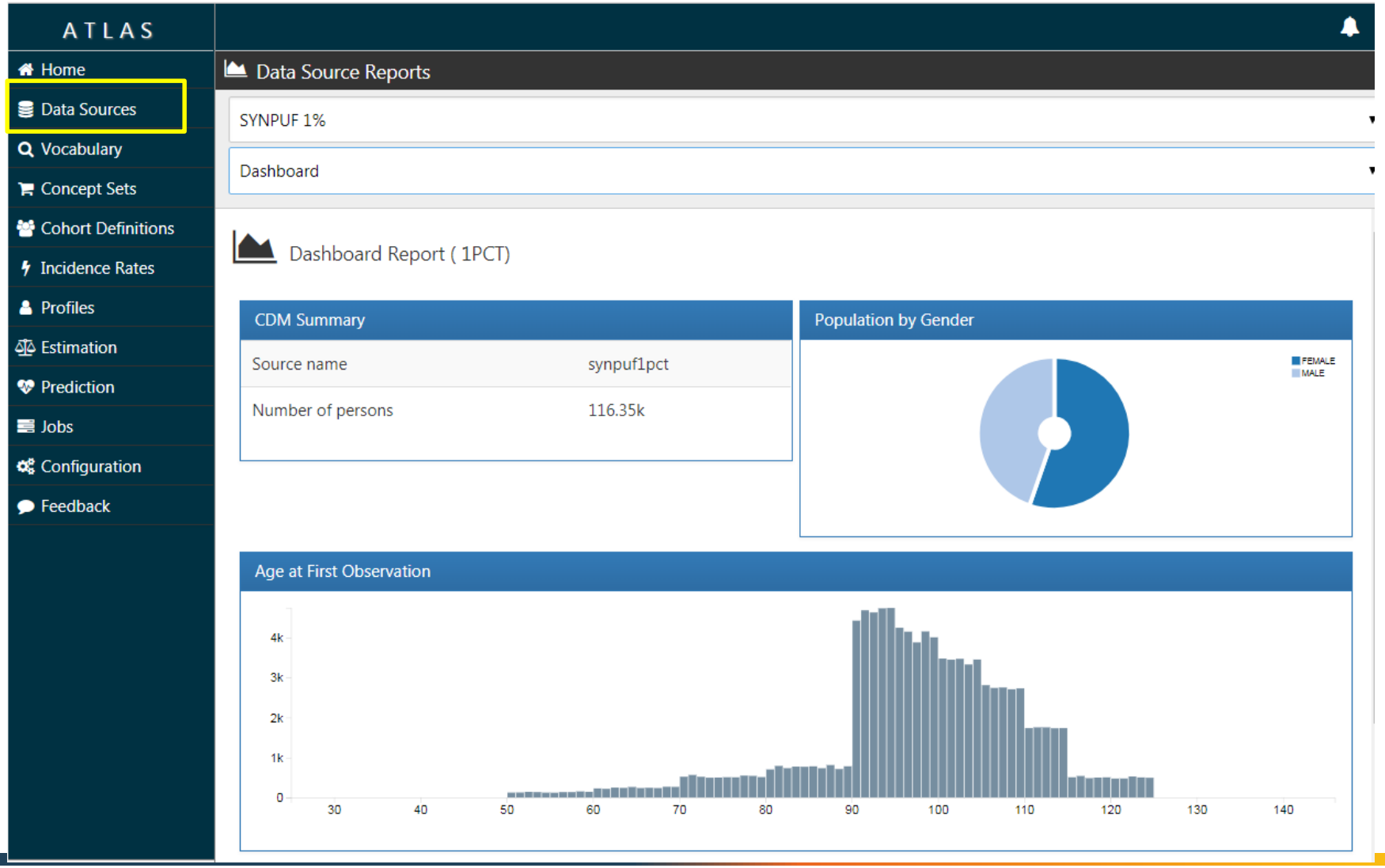
ATLAS 是什么？

- 一个免费网络开源软件工具
- 定义查询病人的条件（比如，近五年
- 内有二型糖尿病但没有高血压）
- 运用查询条件找到符合条件的病人
- 分析病人的特征 (characteristics)
- 对病人特征的描述可分享，重用, 可
- 自动在不同系统间转换JSON, SQL, etc.

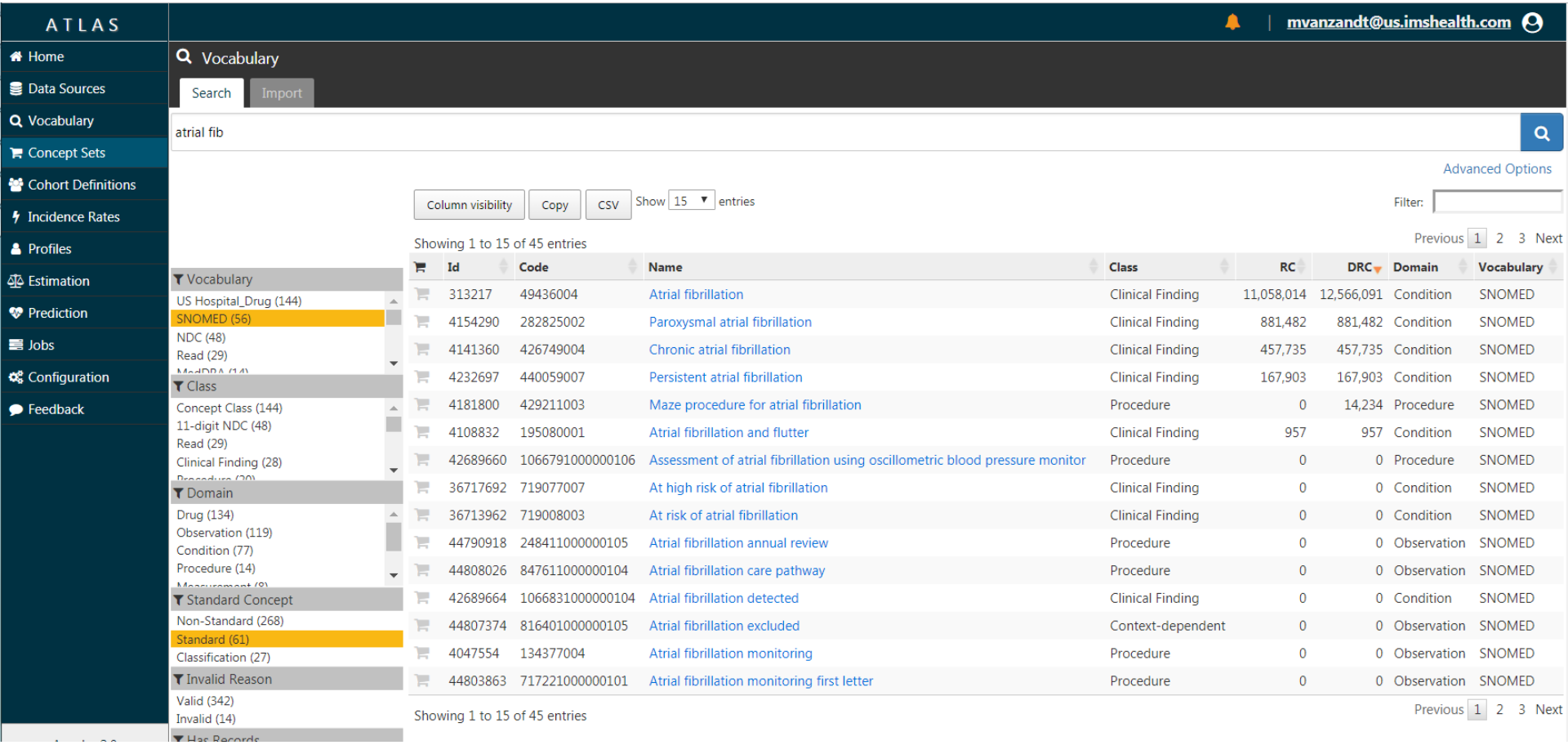
ATLAS 能干什么？

1. 浏览数据源
2. 检索术语Vocabulary
3. 定义概念集Concept Set
4. 定义病人组和他们的临床特征
5. 查询数据库找到符合条件的病人组
6. 可视化单独病人的情况 (Patient profile)
7. 做人群的效果 (results)估计 (Pop)

浏览数据源


ATLAS
Home
Data Sources
Vocabulary
Concept Sets
Cohort Definitions
Incidence Rates
Profiles
Estimation
Prediction
Jobs
Configuration
Feedback
Data Source Reports
SYNPUF 1%
Dashboard
Dashboard Report (1PCT)
CDM Summary
Source name
synpuf1pct
Number of persons
116.35k
Population by Gender
FEMALE
MALE
Age at First Observation
4k
3k
2k
1k
0
30
40
50
60
70
80
90
100
110
120
130
140

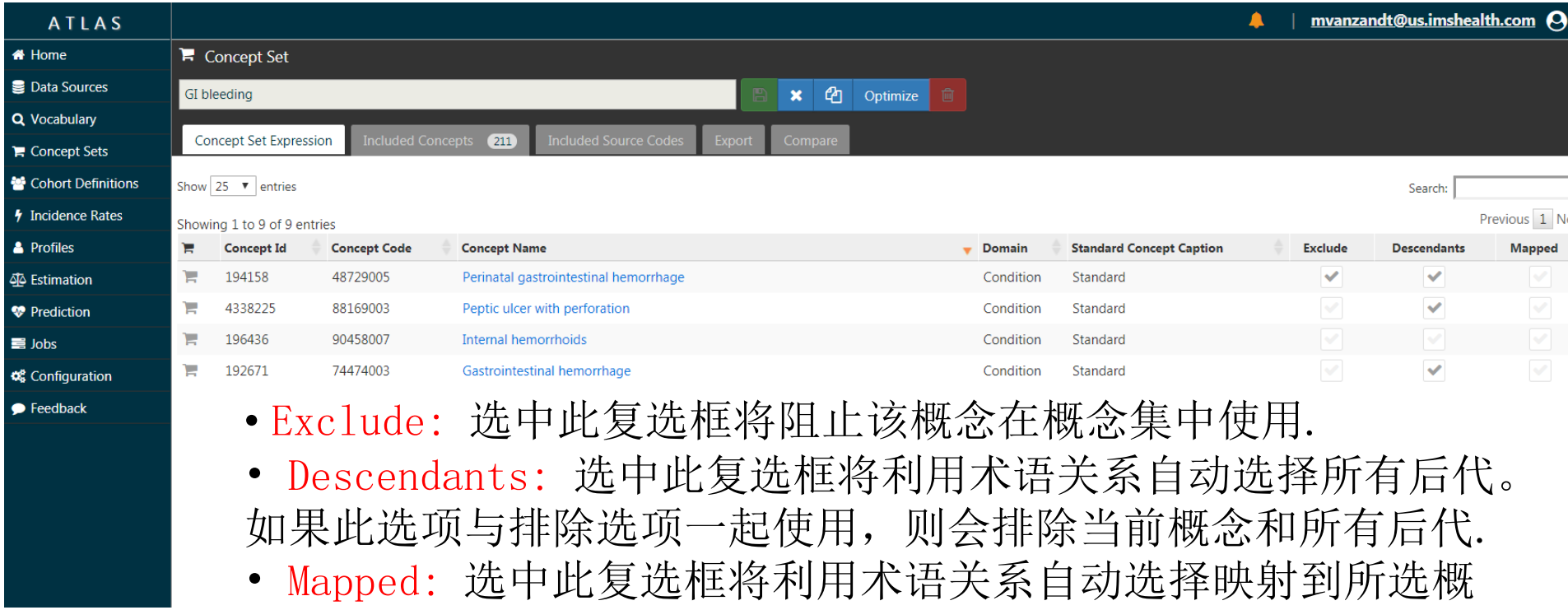
检索术语


ATLAS
mvanzandt@us.imshealth.com
Home
Data Sources
Vocabulary
Concept Sets
Cohort Definitions
Incidence Rates
Profiles
Estimation
Prediction
Jobs
Configuration
Feedback
Vocabulary
Search
Import
atrial fib
Advanced Options
Column visibility
Copy
CSV
Show 15 entries
Filter:
Showing 1 to 15 of 45 entries
Previous 1 2 3 Next
Vocabulary
US Hospital_Drug (144)
SNOMED (56)
NDC (48)
Read (29)
Class
Concept Class (144)
11-digit NDC (48)
Read (29)
Clinical Finding (28)
Domain
Drug (134)
Observation (119)
Condition (77)
Procedure (14)
Standard Concept
Non-Standard (268)
Standard (61)
Classification (27)
Invalid Reason
Valid (342)
Invalid (14)
Id Code Name Class RC DRC Domain Vocabulary
313217 49436004 Atrial fibrillation Clinical Finding 11,058,014 12,566,091 Condition SNOMED
4154290 282825002 Paroxysmal atrial fibrillation Clinical Finding 881,482 881,482 Condition SNOMED
4141360 426749004 Chronic atrial fibrillation Clinical Finding 457,735 457,735 Condition SNOMED
4232697 440059007 Persistent atrial fibrillation Clinical Finding 167,903 167,903 Condition SNOMED
4181800 429211003 Maze procedure for atrial fibrillation Procedure 0 14,234 Procedure SNOMED
4108832 195080001 Atrial fibrillation and flutter Clinical Finding 957 957 Condition SNOMED
42689660 1066791000000106 Assessment of atrial fibrillation using oscillometric blood pressure monitor Procedure 0 0 Procedure SNOMED
36717692 719077007 At high risk of atrial fibrillation Clinical Finding 0 0 Condition SNOMED
36713962 719008003 At risk of atrial fibrillation Clinical Finding 0 0 Condition SNOMED
44790918 248411000000105 Atrial fibrillation annual review Procedure 0 0 Observation SNOMED
44808026 847611000000104 Atrial fibrillation care pathway Procedure 0 0 Observation SNOMED
42689664 1066831000000104 Atrial fibrillation detected Clinical Finding 0 0 Condition SNOMED
44807374 816401000000105 Atrial fibrillation excluded Context-dependent 0 0 Observation SNOMED
4047554 134377004 Atrial fibrillation monitoring Procedure 0 0 Observation SNOMED
44803863 717221000000101 Atrial fibrillation monitoring first letter Procedure 0 0 Observation SNOMED
Showing 1 to 15 of 45 entries
Previous 1 2 3 Next

检索术语

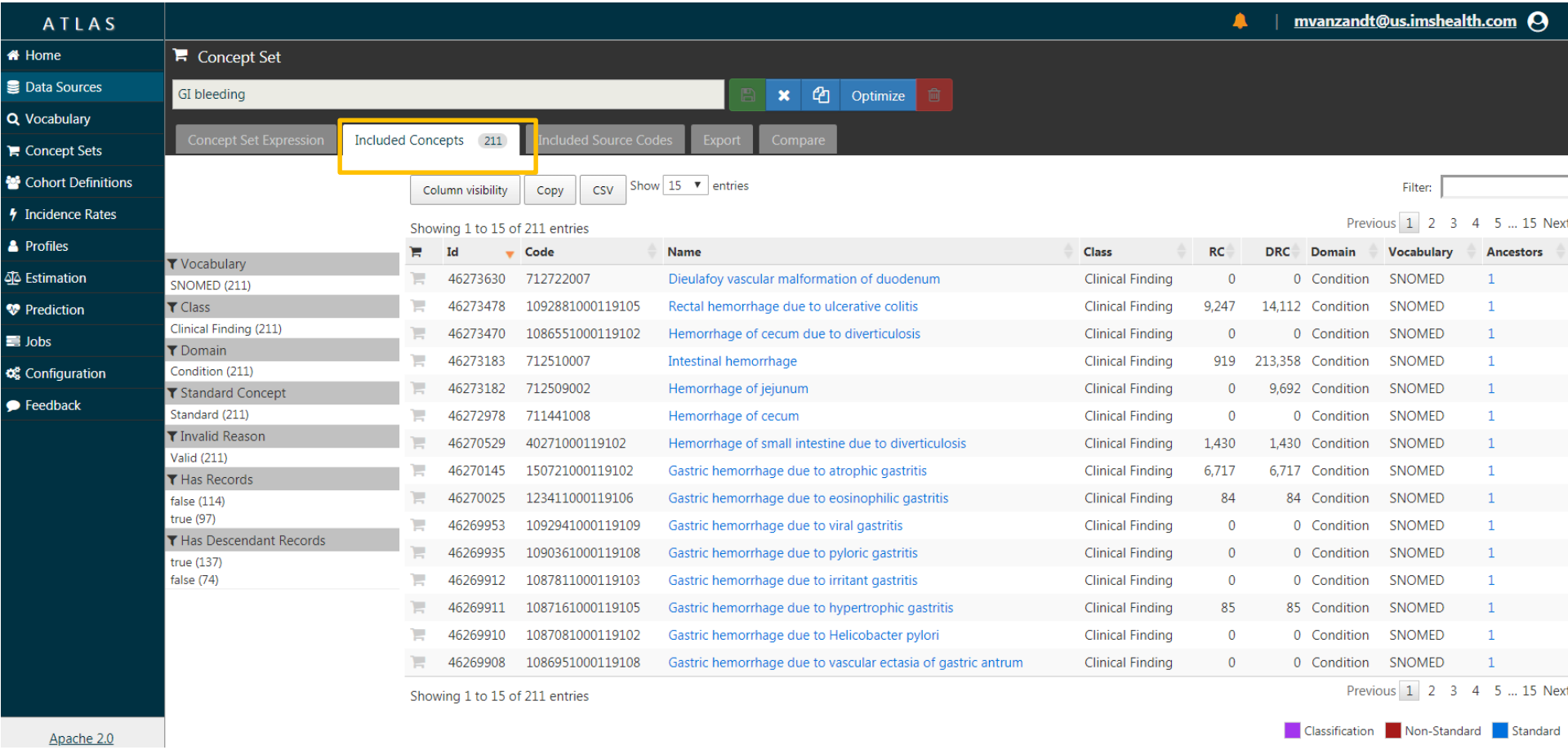
- ID: OHDSI OMOP CDM中一个唯一的concept ID
- Concept Code: 源术语集中的概念识别符
- Class: 由源术语集定义的类别
- RC: 记录数. 显示CDM中用这个概念编码的记录的数量
- DRC: 后代记录数. DRC列将显示在CDM中编码的所有后代概念的总和.
- Domain: OMOP CDM 定义的类别 (e.g.,疾病, 病人, 观察, 样品, etc)

定义概念集 Concept Set

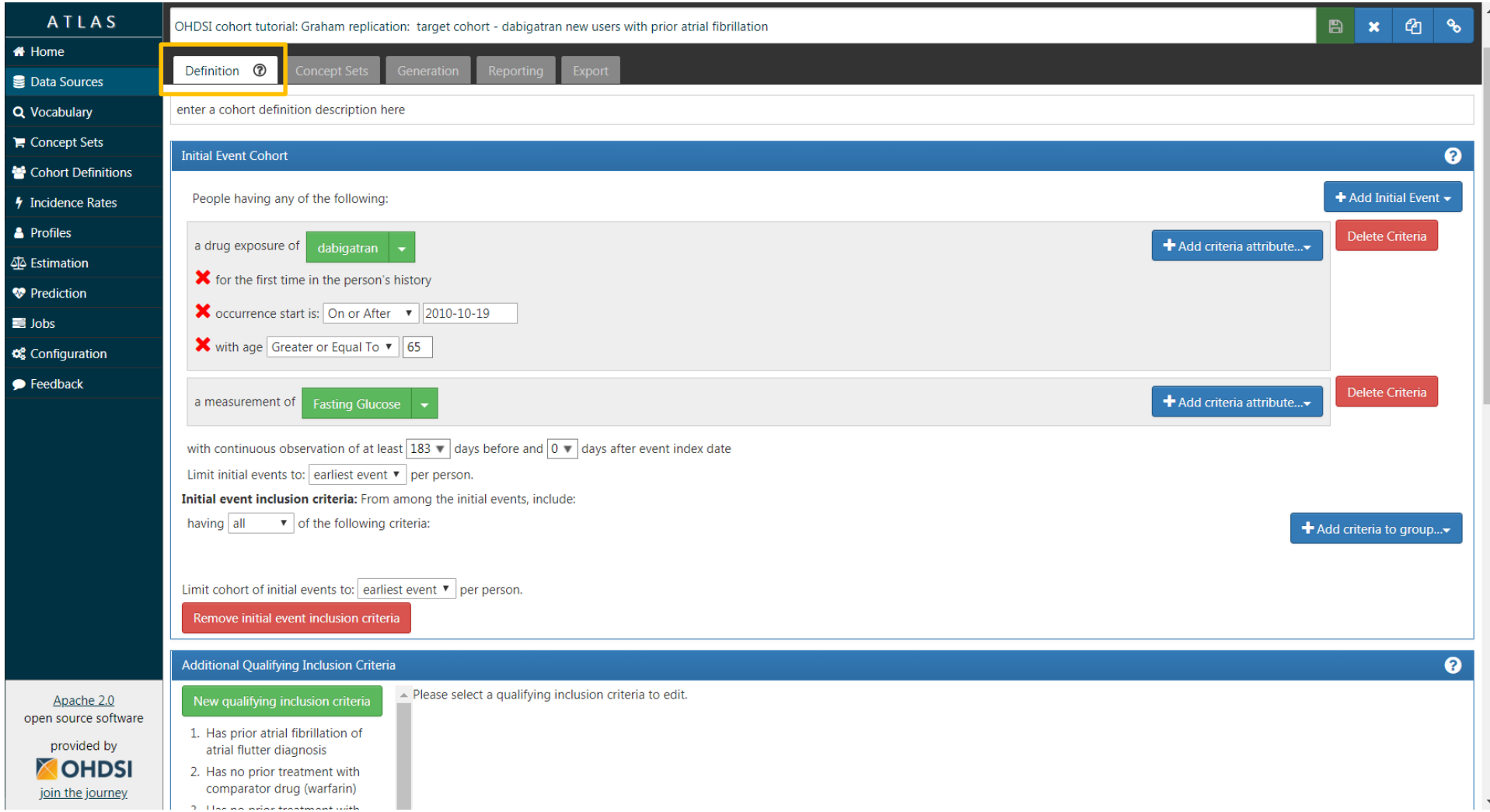


- Exclude：选中此复选框将阻止该概念在概念集中使用.
- Descendants：选中此复选框将利用术语关系自动选择所有后代。如果此选项与排除选项一起使用，则会排除当前概念和所有后代.
- Mapped：选中此复选框将利用术语关系自动选择映射到所选概念的所有概念

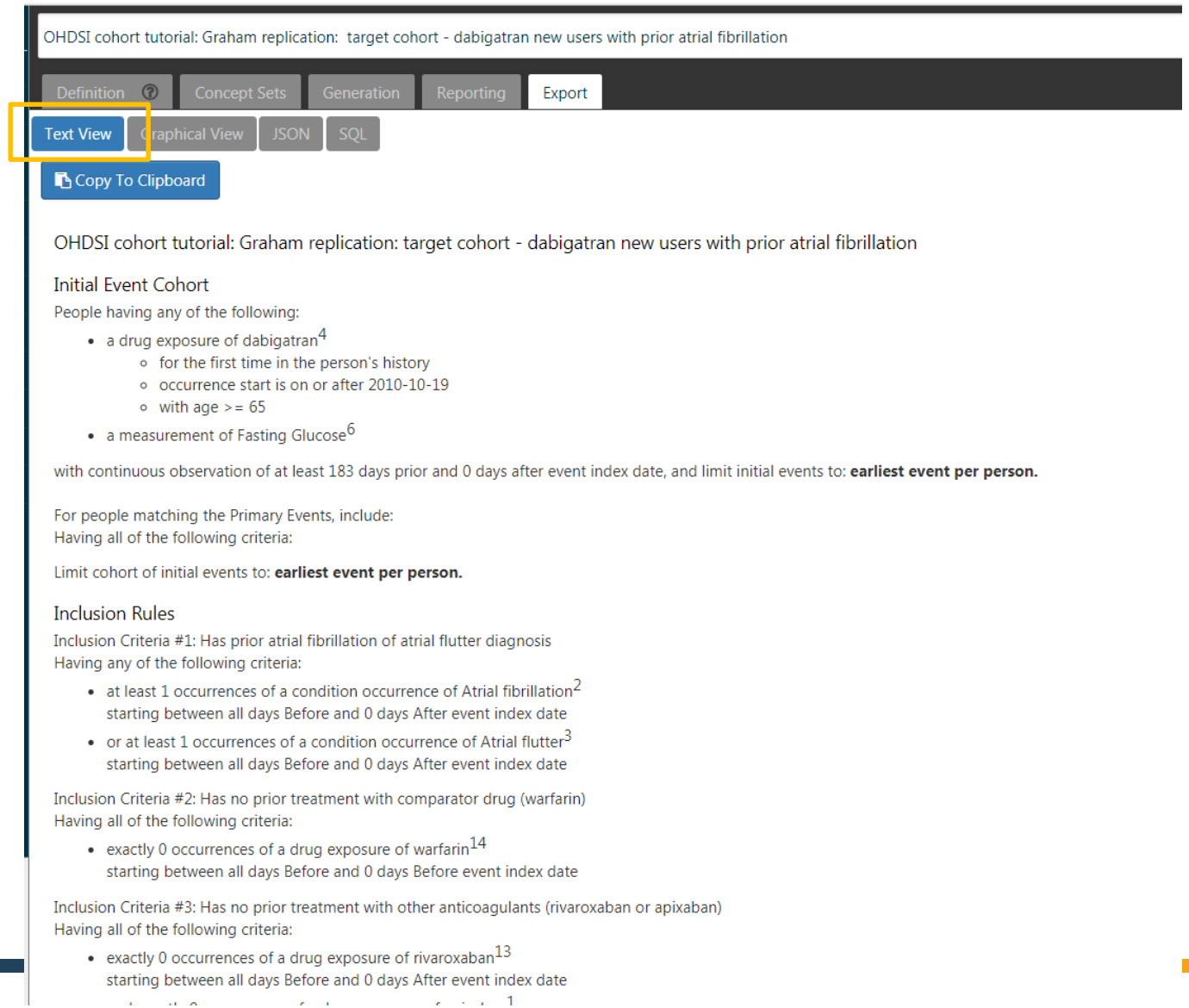
定义概念集


ATLAS
mvanzandt@us.imshealth.com
Home
Data Sources
Vocabulary
Concept Sets
Cohort Definitions
Incidence Rates
Profiles
Estimation
Prediction
Jobs
Configuration
Feedback
Apache 2.0
Concept Set
GI bleeding
Optimize
Concept Set Expression
Included Concepts 211
Included Source Codes
Export
Compare
Column visibility
Copy
CSV
Show 15 entries
Filter:
Showing 1 to 15 of 211 entries
Previous 1 2 3 4 5 … 15 Next
Vocabulary
SNOMED (211)
Class
Clinical Finding (211)
Domain
Condition (211)
Standard Concept
Standard (211)
Invalid Reason
Valid (211)
Has Records
false (114)
true (97)
Has Descendant Records
true (137)
false (74)
Id
Code
Name
Class
RC
DRC
Domain
Vocabulary
Ancestors
46273630 712722007 Dieulafoy vascular malformation of duodenum Clinical Finding 0 0 Condition SNOMED 1
46273478 1092881000119105 Rectal hemorrhage due to ulcerative colitis Clinical Finding 9,247 14,112 Condition SNOMED 1
46273470 1086551000119102 Hemorrhage of cecum due to diverticulosis Clinical Finding 0 0 Condition SNOMED 1
46273183 712510007 Intestinal hemorrhage Clinical Finding 919 213,358 Condition SNOMED 1
46273182 712509002 Hemorrhage of jejunum Clinical Finding 0 9,692 Condition SNOMED 1
46272978 711441008 Hemorrhage of cecum Clinical Finding 0 0 Condition SNOMED 1
46270529 40271000119102 Hemorrhage of small intestine due to diverticulosis Clinical Finding 1,430 1,430 Condition SNOMED 1
46270145 150721000119102 Gastric hemorrhage due to atrophic gastritis Clinical Finding 6,717 6,717 Condition SNOMED 1
46270025 123411000119106 Gastric hemorrhage due to eosinophilic gastritis Clinical Finding 84 84 Condition SNOMED 1
46269953 1092941000119109 Gastric hemorrhage due to viral gastritis Clinical Finding 0 0 Condition SNOMED 1
46269935 1090361000119108 Gastric hemorrhage due to pyloric gastritis Clinical Finding 0 0 Condition SNOMED 1
46269912 1087811000119103 Gastric hemorrhage due to irritant gastritis Clinical Finding 0 0 Condition SNOMED 1
46269911 1087161000119105 Gastric hemorrhage due to hypertrophic gastritis Clinical Finding 85 85 Condition SNOMED 1
46269910 1087081000119102 Gastric hemorrhage due to Helicobacter pylori Clinical Finding 0 0 Condition SNOMED 1
46269908 1086951000119108 Gastric hemorrhage due to vascular ectasia of gastric antrum Clinical Finding 0 0 Condition SNOMED 1
Showing 1 to 15 of 211 entries
Previous 1 2 3 4 5 … 15 Next
Classification
Non-Standard
Standard

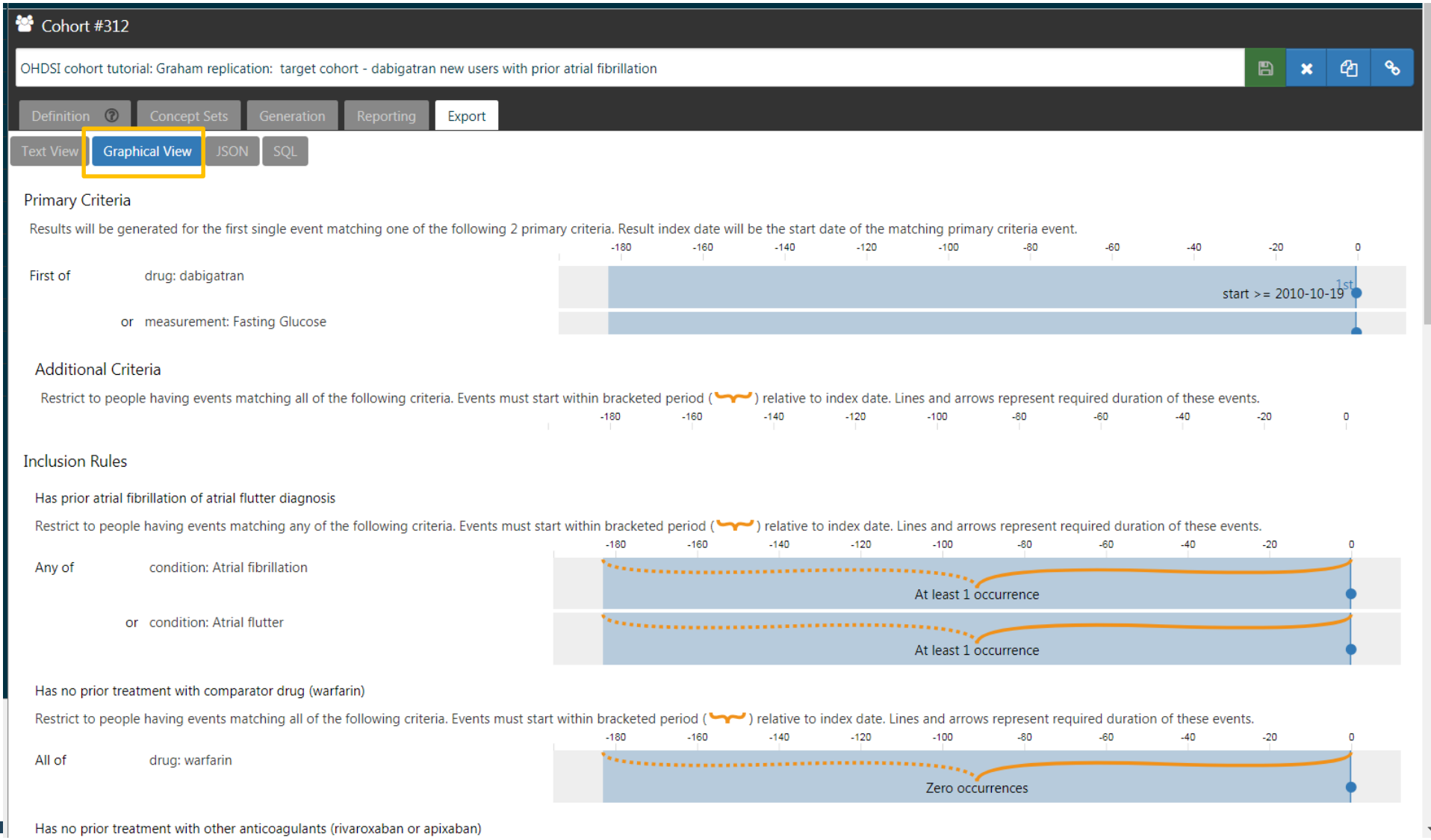
定义病人队列和他们的临床特征



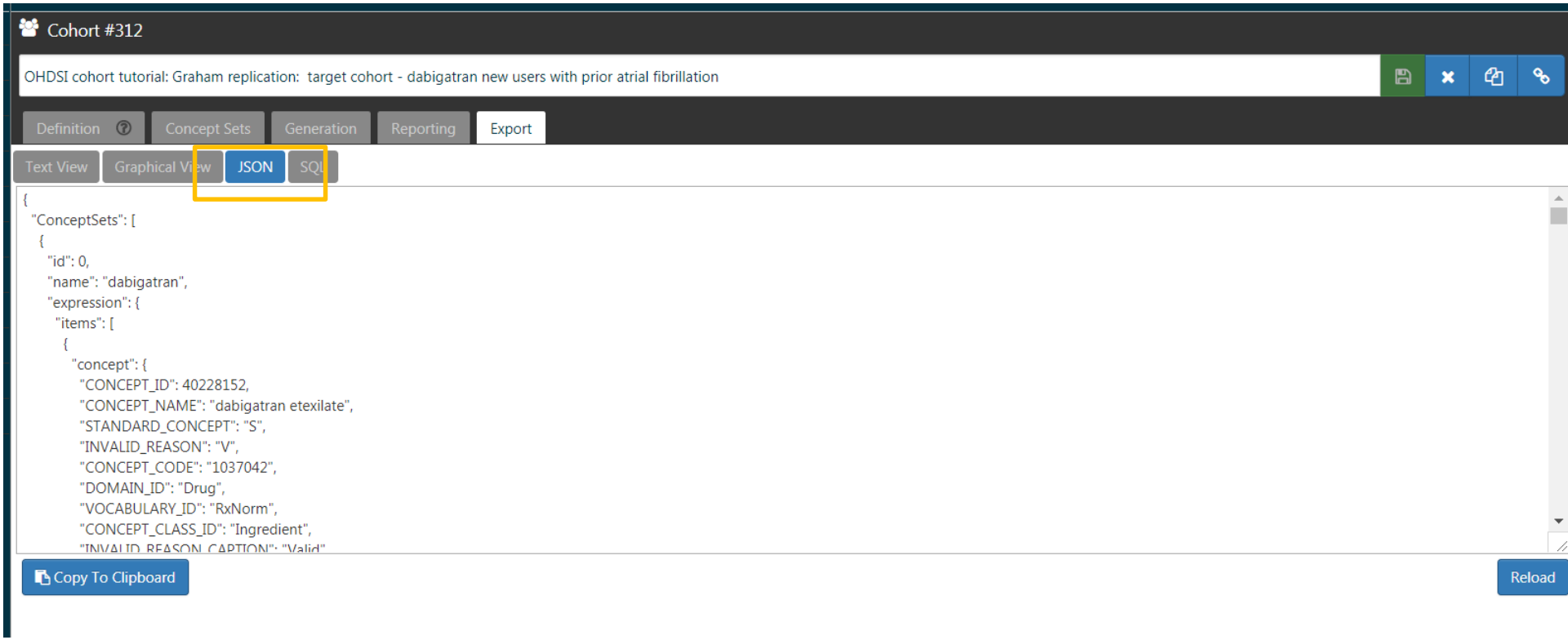
可分享的人群队列定义


OHDSI cohort tutorial: Graham replication: target cohort - dabigatran new users with prior atrial fibrillation
Definition
Concept Sets
Generation
Reporting
Export
Text View
Graphical View
JSON
SQL
Copy To Clipboard
OHDSI cohort tutorial: Graham replication: target cohort - dabigatran new users with prior atrial fibrillation
Initial Event Cohort
People having any of the following:
a drug exposure of dabigatran[4]
for the first time in the person's history
occurrence start is on or after 2010-10-19
with age >= 65
a measurement of Fasting Glucose[6]
with continuous observation of at least 183 days prior and 0 days after event index date, and limit initial events to: earliest event per person.
For people matching the Primary Events, include:
Having all of the following criteria:
Limit cohort of initial events to: earliest event per person.
Inclusion Rules
Inclusion Criteria #1: Has prior atrial fibrillation of atrial flutter diagnosis
Having any of the following criteria:
at least 1 occurrences of a condition occurrence of Atrial fibrillation[2]
starting between all days Before and 0 days After event index date
or at least 1 occurrences of a condition occurrence of Atrial flutter[3]
starting between all days Before and 0 days After event index date
Inclusion Criteria #2: Has no prior treatment with comparator drug (warfarin)
Having all of the following criteria:
exactly 0 occurrences of a drug exposure of warfarin[14]
starting between all days Before and 0 days Before event index date
Inclusion Criteria #3: Has no prior treatment with other anticoagulants (rivaroxaban or apixaban)
Having all of the following criteria:
exactly 0 occurrences of a drug exposure of rivaroxaban[13]
starting between all days Before and 0 days After event index date

可分享的人群队列定义


Cohort #312
OHDSI cohort tutorial: Graham replication: target cohort - dabigatran new users with prior atrial fibrillation
Definition
Concept Sets
Generation
Reporting
Export
Text View
Graphical View
JSON
SQL
Primary Criteria
Results will be generated for the first single event matching one of the following 2 primary criteria. Result index date will be the start date of the matching primary criteria event.
-180 -160 -140 -120 -100 -80 -60 -40 -20 0
First of drug: dabigatran
start >= 2010-10-19
1st
or measurement: Fasting Glucose
Additional Criteria
Restrict to people having events matching all of the following criteria. Events must start within bracketed period ( ) relative to index date. Lines and arrows represent required duration of these events.
-180 -160 -140 -120 -100 -80 -60 -40 -20 0
Inclusion Rules
Has prior atrial fibrillation of atrial flutter diagnosis
Restrict to people having events matching any of the following criteria. Events must start within bracketed period ( ) relative to index date. Lines and arrows represent required duration of these events.
-180 -160 -140 -120 -100 -80 -60 -40 -20 0
Any of condition: Atrial fibrillation
At least 1 occurrence
or condition: Atrial flutter
At least 1 occurrence
Has no prior treatment with comparator drug (warfarin)
Restrict to people having events matching all of the following criteria. Events must start within bracketed period ( ) relative to index date. Lines and arrows represent required duration of these events.
-180 -160 -140 -120 -100 -80 -60 -40 -20 0
All of drug: warfarin
Zero occurrences
Has no prior treatment with other anticoagulants (rivaroxaban or apixaban)

可分享的人群队列定义


Cohort #312
OHDSI cohort tutorial: Graham replication: target cohort - dabigatran new users with prior atrial fibrillation
Definition
Concept Sets
Generation
Reporting
Export
Text View
Graphical View
JSON
SQL
{
"ConceptSets": [
{
"id": 0,
"name": "dabigatran",
"expression": {
"items": [
{
"concept": {
"CONCEPT_ID": 40228152,
"CONCEPT_NAME": "dabigatran etexilate",
"STANDARD_CONCEPT": "S",
"INVALID_REASON": "V",
"CONCEPT_CODE": "1037042",
"DOMAIN_ID": "Drug",
"VOCABULARY_ID": "RxNorm",
"CONCEPT_CLASS_ID": "Ingredient",
Copy To Clipboard
Reload

可分享的人群队列定义

Cohort #312

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

Definition | Concept Sets | Generation | Reporting | Export

Text View | Graphical View | JSON | SQL

Template OHDSI.SQL | MSSQL Server | MS APS | Oracle | PostgreSQL | Amazon Red Shift | Impala | Netezza

Copy To Clipboard

```
CREATE TABLE #Codesets (
  codeset_id int NOT NULL,
  concept_id bigint NOT NULL
)
;

INSERT INTO #Codesets (codeset_id, concept_id)
SELECT 0 as codeset_id, c.concept_id FROM (select distinct I.concept_id FROM
(
  select concept_id from @vocabulary_database_schema.CONCEPT where concept_id in (40228152)and invalid_reason is null
UNION  select c.concept_id
  from @vocabulary_database_schema.CONCEPT c
  join @vocabulary_database_schema.CONCEPT_ANCESTOR ca on c.concept_id = ca.descendant_concept_id
  and ca.ancestor_concept_id in (40228152)
  and c.invalid_reason is null

) I
) C;
INSERT INTO #Codesets (codeset_id, concept_id)
SELECT 1 as codeset_id, c.concept_id FROM (select distinct I.concept_id FROM
(
  select concept_id from @vocabulary_database_schema.CONCEPT where concept_id in (313217)and invalid_reason is null
UNION  select c.concept_id
  from @vocabulary_database_schema.CONCEPT c
  join @vocabulary_database_schema.CONCEPT_ANCESTOR ca on c.concept_id = ca.descendant_concept_id
  and ca.ancestor_concept_id in (313217)
  and c.invalid_reason is null

) I
) C;
```

查询数据库找到符合条件的病人队列

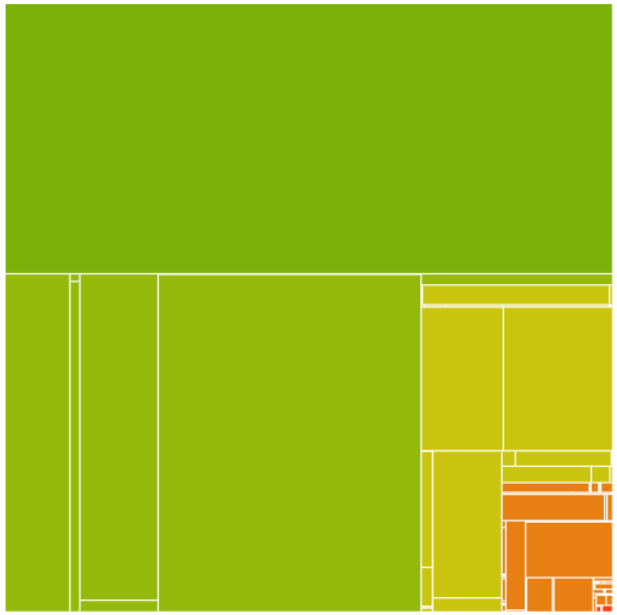
Inclusion Report | Cohort Features

Inclusion Report for Open Claims. Full 2018/01

		Match Rate	Matches	Total
	Summary Statistics:	44.42%	239,580	539,343

	Inclusion Rule	N	% Satisfied	% To-Gain
1.	Has prior atrial fibrillation of atrial flutter diagnosis	348,221	64.56%	24.00%
2.	Has no prior treatment with comparator drug (warfarin)	454,958	84.35%	5.95%
3.	Has no prior treatment with other anticoagulants (rivaroxaban or apixaban)	440,623	81.70%	6.91%
4.	Not in a skilled nursing facility or nursing home, or receiving hospice care on the index date	538,670	99.88%	0.04%
5.	Not undergoing dialysis or kidney transplant recipient	538,977	99.93%	0.02%
6.	No mitral valve disease, heart valve repair, or replacement in the prior 6 months	526,632	97.64%	0.88%
7.	No deep vein thrombosis or pulmonary embolism in the prior 6 months	515,782	95.63%	0.58%
8.	No joint replacement surgery in the prior 6 months	535,700	99.32%	0.25%

Population Visualization Switch to attrition view



练习

- 创建关注组，对照组，以及终点事件队列

关注组Target Cohort

满足下列任何条件的病人 People having any of the following:

- 首次使用达比加群（ dabigatran）的患者
  - 首次使用日期晚于 2010-10-19
  - 年龄 >= 65
- 一次空腹血糖测定记录
- 在事件索引日期前至少183天和0天后连续观察，并将初始事件限制为：每人最早事件。

同时，满足以下条件:

入组标准

入组标准#1:

至少发生一次（at least once）房颤诊断

- 在事件索引日期之前所有时间（all days）至之后的0天（0 day）

或者至少发生一次（at least once）房扑诊断

- 在事件索引日期之前所有时间（all days）至之后的0天（0 day）

- 入组表准#2: 没有使用过华法林（exact 0 occurrence）
  - 在事件索引日期之前所有时间（all days）至之前的0天（0 day）

确定观察结束时间： 最后一次使用达比加群的日期

- 每次暴露指间允许3天的间隔
- 暴露结束后额外加0天

对照组Comparator Cohort

满足下列任何条件的病人 People having any of the following:

•达比加群（dabigatran）使用者

–首次使用

–晚于 2010-10-19

–年龄 >= 65

•在事件索引日期前至少183天和0天后连续观察，并将初始事件限制为：每人最早事件。

•将合格队列限制为：每人最早事件。

确定观察结束时间： 最后一次使用达比加群的日期

•每次暴露指间允许3天的间隔

•暴露结束后额外加0天

终点事件队列Outcome Cohort

满足下列任何条件的病人 People having any of the following:

- 一次缺血性卒中的诊断
  - –诊断类型包括任何一种：住院病人详单-主要诊断inpatient detail - primary，住院病人抬头-主要诊断Inpatient header - primary，主要诊断Primary Condition，住院病人详单-首位诊断Inpatient detail - 1st position，住院病人抬头-首位诊断Inpatient header - 1st position
  - –就诊属于任何一种：急诊，住院
- 在事件索引日期前至少183天开始和0天后连续观察，并将初始事件限制为：每人最早事件。
- 将合格队列限制为：每人最早事件。

附加条件

最多发生0次缺血性卒中的诊断（at most 0 occurrence）

- –在索引时间之前30天和之前1天之间

将初始事件限制为：每人最早事件。

将合格队列限制为：每人最早事件。

确定观察结束时间

不选择观察结束时间. 默认设定下，队列结束时间是包含索引事件的观察期结束时间


OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

人群估计

用Atlas确定队列限定条件

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

Poisson regression

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

0 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:

0

Remove patients who enter both cohorts?

Yes

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

No

用ATLAS选择协变量

Use propensity score adjustment as a confounding adjustment strategy for baseline covariates?

Yes ▼

Which types of baseline covariates do you want to include in the propensity score model?

- ☑ Demographics
  - ☑ Gender
  - ☑ Age group (5-year bands)
  - ☑ Index year
  - ☐ Index month
  - ☐ Race
  - ☐ Ethnicity
- ☑ Conditions
  - ☐ In prior 30d
  - ☑ In prior 365d
  - ☐ In prior 180d within inpatient setting
  - ☐ All time prior
  - ☐ Overlapping index date
- ☐ Condition aggregation
  - ☐ SNOMED
  - ☐ MedDRA
- ☑ Drugs
  - ☐ In prior 30d
  - ☐ In prior 365d
  - ☐ All time prior
  - ☑ Overlapping index date
- ☑ Drug aggregation
  - ☐ Clinical Drug
  - ☑ Ingredient
  - ☑ ATC Class
- ☐ Procedures
  - ☐ In prior 30d
  - ☐ In prior 365d
- ☐ Measurement
  - ☐ Existence in prior 30d
  - ☐ Existence in prior 365d
  - ☐ Count in prior 365d
  - ☐ Has latest prior numeric value below normal range
  - ☐ Has latest prior numeric value above normal range
- ☐ Risk scores
  - ☐ Charlson
  - ☐ CHADS2
  - ☐ DCSI
- ☑ Concept counts (count of distinct conditions/procedures/visits in history)
- ☐ Interaction terms
  - ☐ By index year
  - ☐ By index month

What concepts do you want to include in baseline covariates in the propensity score model? (Leave blank if you want to include everything)

OHDSI estimation tutorial- Garbe replication: covariates to include in PS model

What concepts do you want to exclude from baseline covariates in the propensity score model? (Leave blank if you want to include everything)

OHDSI estimation tutorial- Garbe replication: covariates to exclude in PS model

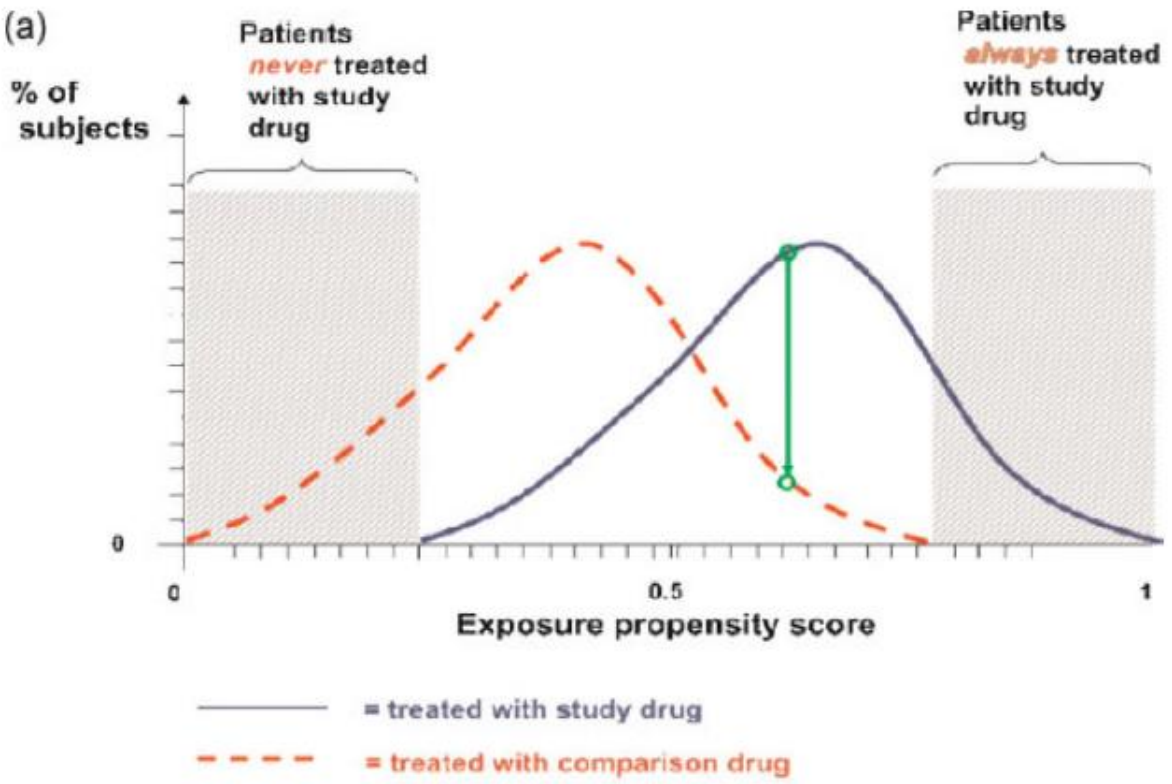
用ATLAS控制协变量

Use propensity score adjustment as a confounding adjustment strategy for baseline covariates?

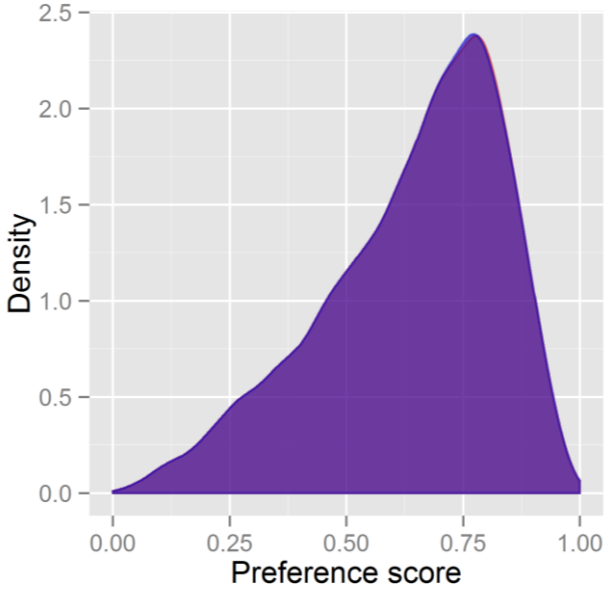
No ▼

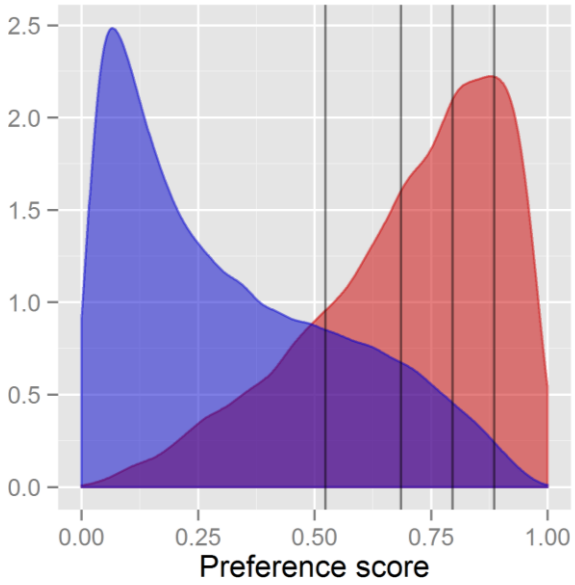
Do you want to adjust for baseline covariates in the outcome model?

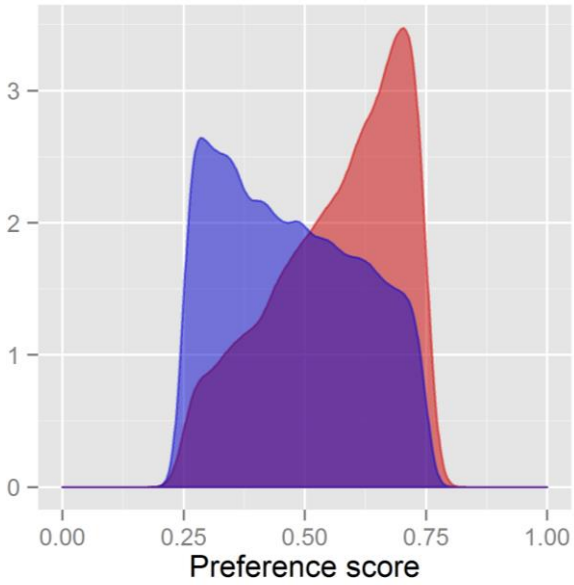
No ▼



用propensity score控制混杂因素







Matching
For every treated subject, select n comparators using greedy matching

Stratification
Stratify into equally-sized strata based on PS

Trimming
Remove subjects with high and low PS

标准化方法库

Estimation methods

Cohort Method
使用大规模回归进行倾向（propensity）s 和结果（outcome）模型的新用户队列研究

Self-Controlled Case Serie
使用少量或者较多预测变量的自身对照病例系列分析，包括年龄和季节性回归样条s

Self-Controlled Cohort
一种自身对照的队列研究设计，使用暴露之前的时间段作为对照组s

IC Temporal Pattern Disc.
一种自身对照的设计，但使用其他暴露和结果周围的时间模式来纠正时变混杂。s

Case-control
病例-对照研究，使用年龄，性别，医疗提供者，以及就诊时间来匹配对照。允许将研究嵌套在另一个队列中s

Prediction methods

Patient Level Prediction
使用多种机器学习算法构建和评估用户个体化终点事件的预测模型。

Feature Extraction
使用CDM中的数据为用户创建的队列自动提取大量的特征值

Method characterization

Empirical Calibration
使用阴性对照暴露 - 结果配对来分析和校准特定的分析设计

Method Evaluation
使用实际数据，建立的参考集，以及混合实际数据的模拟数据来评估方法的性能。

Supporting packages

Database Connector
直接连接到各种数据库平台，包括SQL Server，Oracle和PostgreSQL。

Sql Render
为各种不同格式的SQL语言生成SQL语句

Cyclops
高效运行正则化的logistic, Poisson 和 Cox 回归.

Ohdsi R Tools
一些不能归类为其他类别的支持性工具，包括R数据库的维护工具

Under construction

NULL 分布的阴性对照选择

What outcomes would you like to use as your negative controls? These are concepts known not to be associated with either the target or comparator group, such that we can assume the true relative risk should equal 1. These negative control outcomes will be used for empirical calibration.

OHDSI estimation tutorial: Garbe et al. replication, negative controls

Concept Set

OHDSI estimation tutorial: Garbe et al. replication, negative controls

Save Close

Concept Set Expression | Included Concepts 34 | Included Source Codes | Export

Vocabulary
SNOMED (34)

Class
Clinical Finding (34)

Domain
Condition (34)

Standard Concept
Standard (34)

Invalid Reason
Valid (34)

Has Records
true (34)

Has Descendant Records
true (34)

Show All entries Filter:

Showing 1 to 34 of 34 entries Previous 1 Next

Id	Code	Name	Class	RC	DRC	Domain	Vocabulary
257007	61582004	Allergic rhinitis	Clinical Finding	35,111,076	55,582,634	Condition	SNOMED
255573	13645005	Chronic obstructive lung disease	Clinical Finding	22,235,402	28,534,157	Condition	SNOMED
319843	11851006	Mitral valve disorder	Clinical Finding	4,805,521	6,232,915	Condition	SNOMED
141932	398838000	Seborrheic keratosis	Clinical Finding	4,758,455	4,758,455	Condition	SNOMED
380094	57406009	Carpal tunnel syndrome	Clinical Finding	4,713,600	4,713,600	Condition	SNOMED
313459	73430006	Sleep apnea	Clinical Finding	4,280,415	17,180,747	Condition	SNOMED
436665	13746004	Bipolar disorder	Clinical Finding	3,472,356	7,920,853	Condition	SNOMED
443800	14302001	Amenorrhea	Clinical Finding	3,221,577	3,257,684	Condition	SNOMED
197236	95315005	Uterine leiomyoma	Clinical Finding	3,021,864	3,898,566	Condition	SNOMED
321596	20696009	Peripheral venous insufficiency	Clinical Finding	2,950,514	3,004,234	Condition	SNOMED
436676	47505003	Posttraumatic stress disorder	Clinical Finding	2,933,195	2,978,063	Condition	SNOMED
139099	400097005	Ingrowing nail	Clinical Finding	2,928,631	2,928,631	Condition	SNOMED
314658	8186001	Cardiomegaly	Clinical Finding	2,651,255	2,651,255	Condition	SNOMED
433753	15167005	Alcohol abuse	Clinical Finding	1,758,133	2,013,798	Condition	SNOMED
440374	191736004	Obsessive-compulsive disorder	Clinical Finding	1,666,847	1,666,981	Condition	SNOMED

Revisiting clopidogrel & GI bleed (Opatrny, 2008)

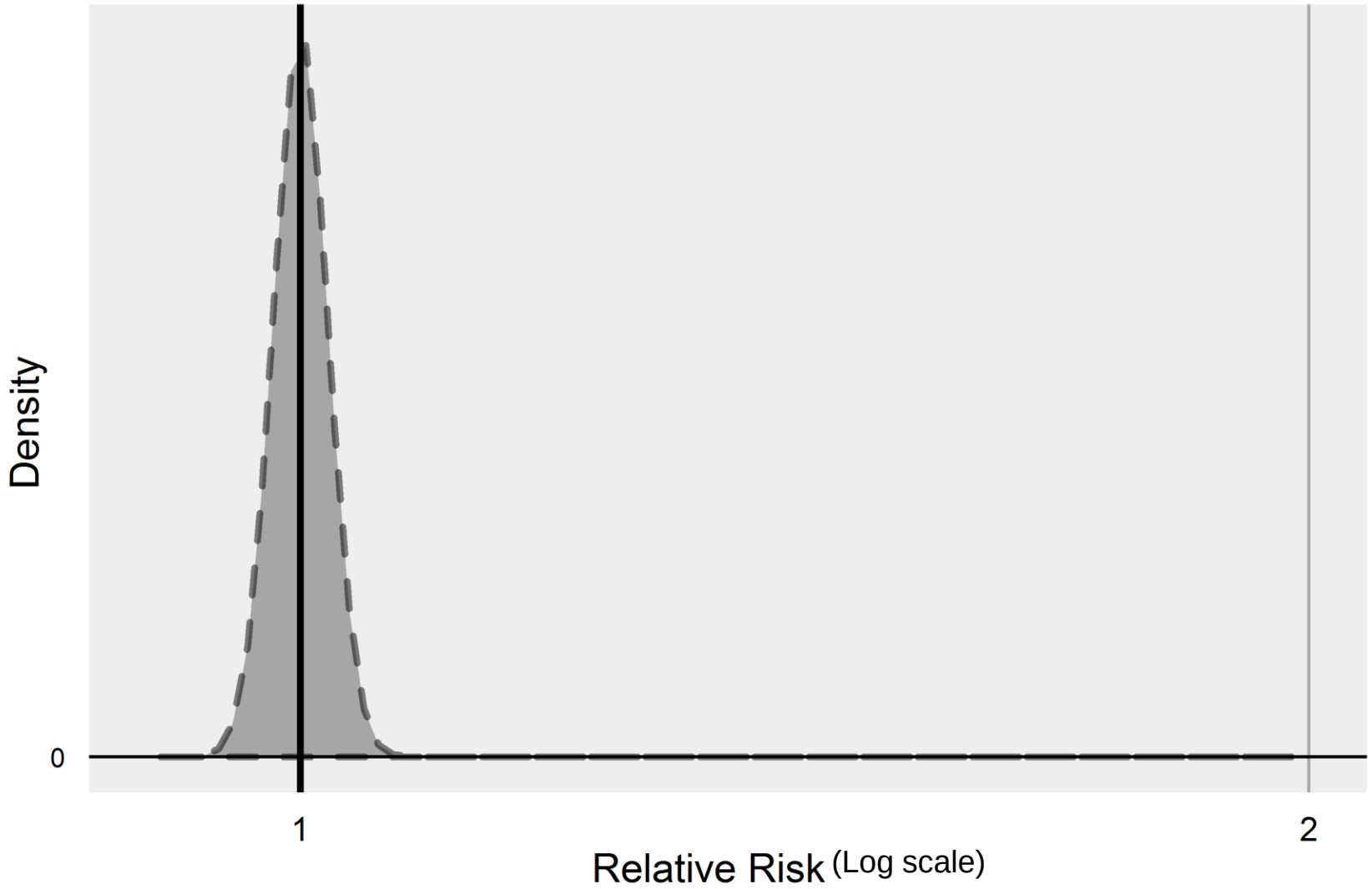
Agent	Cases ($n = 4028$)	Controls ($n = 40\,171$)	Crude rate ratio	Adjusted rate ratio*	95% confidence interval
Antidepressants					
SSRI	335 (8.3%)	1780 (4.4%)	1.97	1.33	1.09, 1.62
TCA	262 (6.5%)	1764 (4.4%)	1.52	1.04	0.83, 1.30
Venlafaxine	56 (1.4%)	229 (0.6%)	2.48	1.85	1.34, 2.55
Anticoagulant					
Warfarin	281 (7.0%)	1130 (2.8%)	2.64	2.17	1.82, 2.59
Clopidogrel	160 (4.0%)	532 (1.3%)	3.16	2.07	1.66, 2.58

OMOP, 2012 (CC: 2000314, CCAE, GI Bleed)

Relative risk: 1.86, 95% CI: 1.79 – 1.93
Standard error: 0.02, p-value: <.001

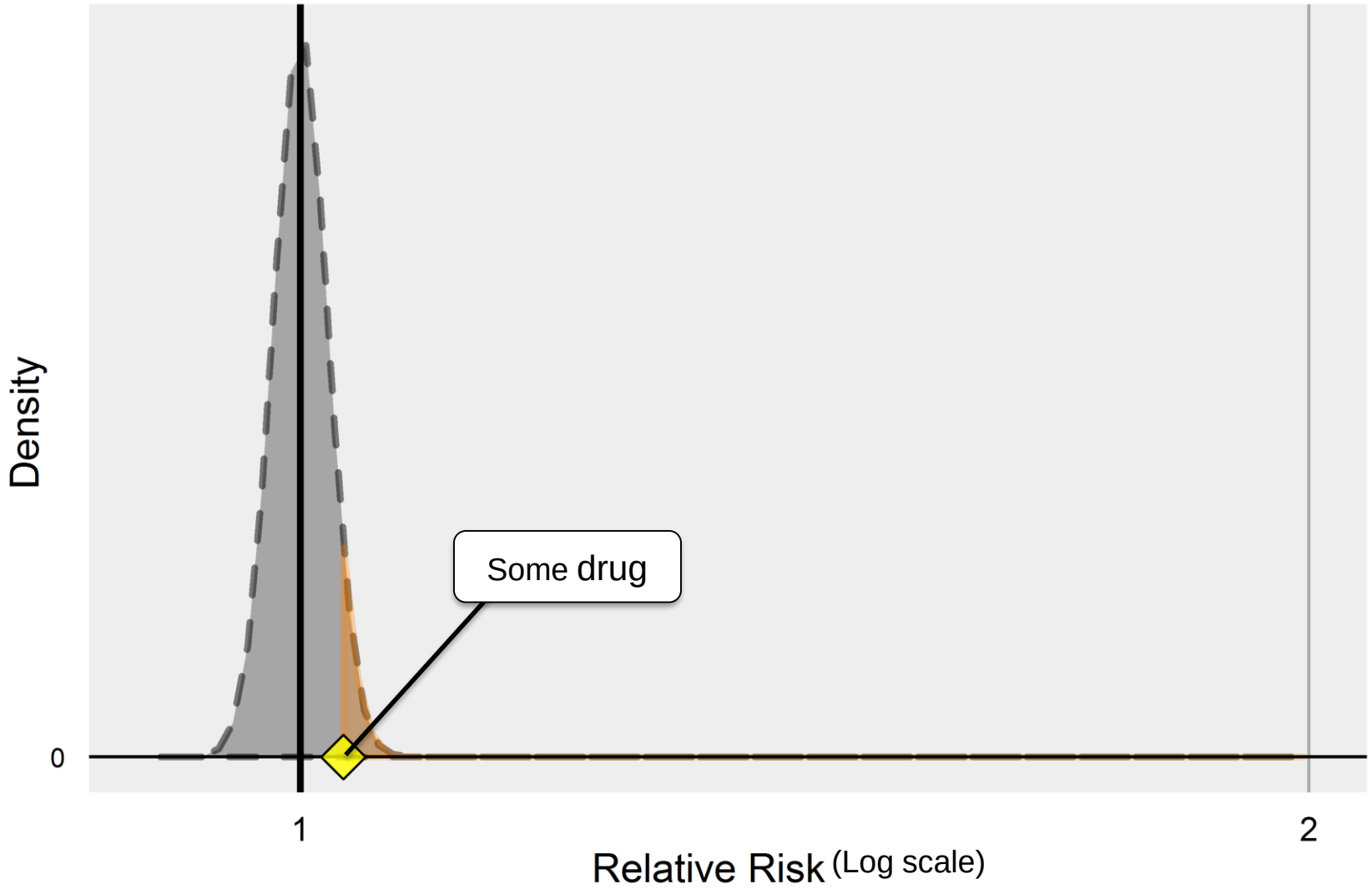
Null 分布

CC: 2000314, CCAE, GI Bleed



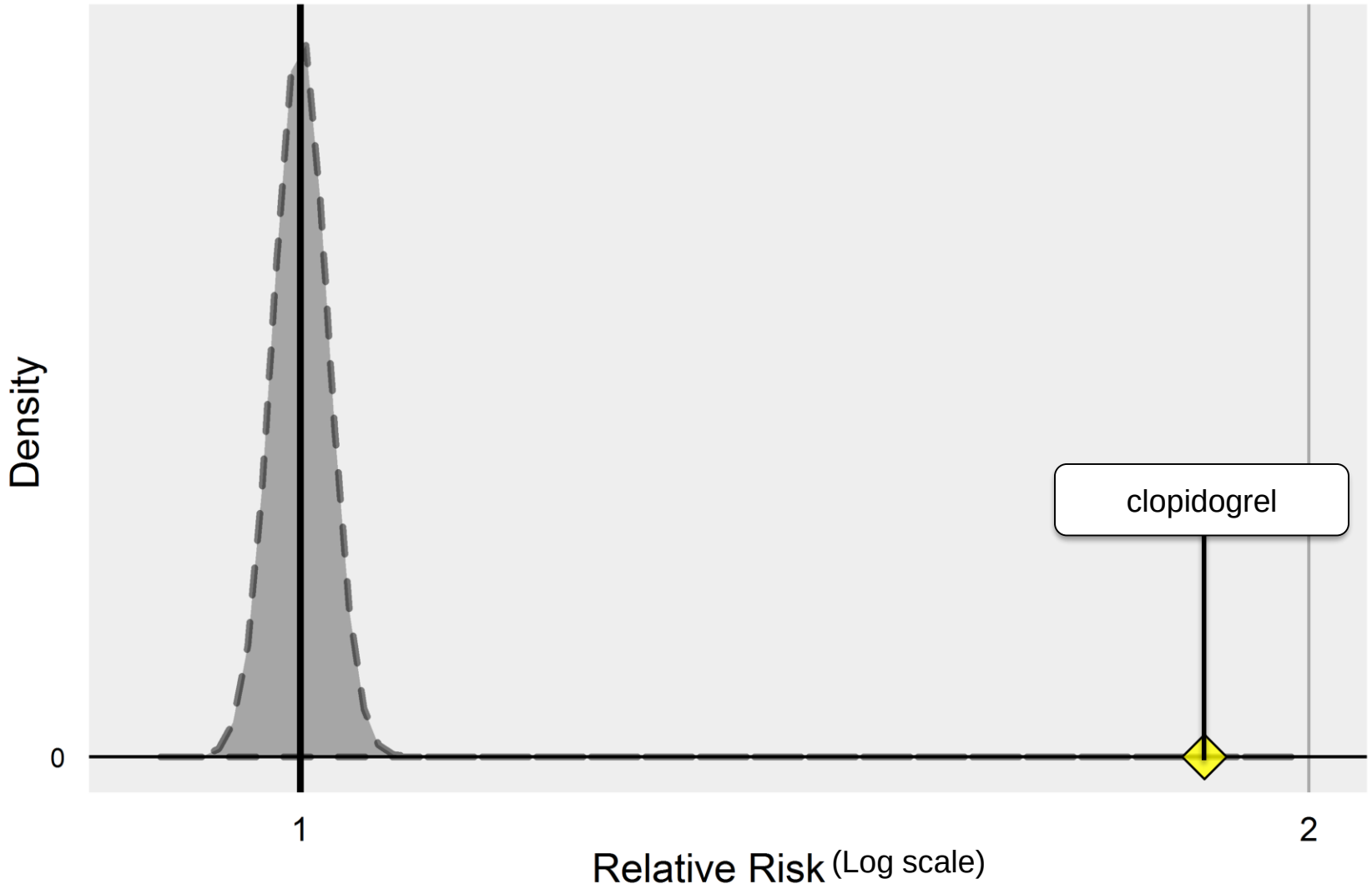
Null 分布

CC: 2000314, CCAE, GI Bleed



Null 分布

CC: 2000314, CCAE, GI Bleed



评价null 分布?

- 现有的p-value计算是假设使用的是无偏估计量(混杂不存在或者被完全纠正)
- 传统做法下, 当 p<.05 时我们拒绝null 假设, 是因为我们认为在这个情况下我们错误拒绝null假设的概率小于0.05. 在观察性研究中这个假设是否成立?
- 我们可以使用阴性对照来测试

OMOP 2011/2012 实验的参考标准

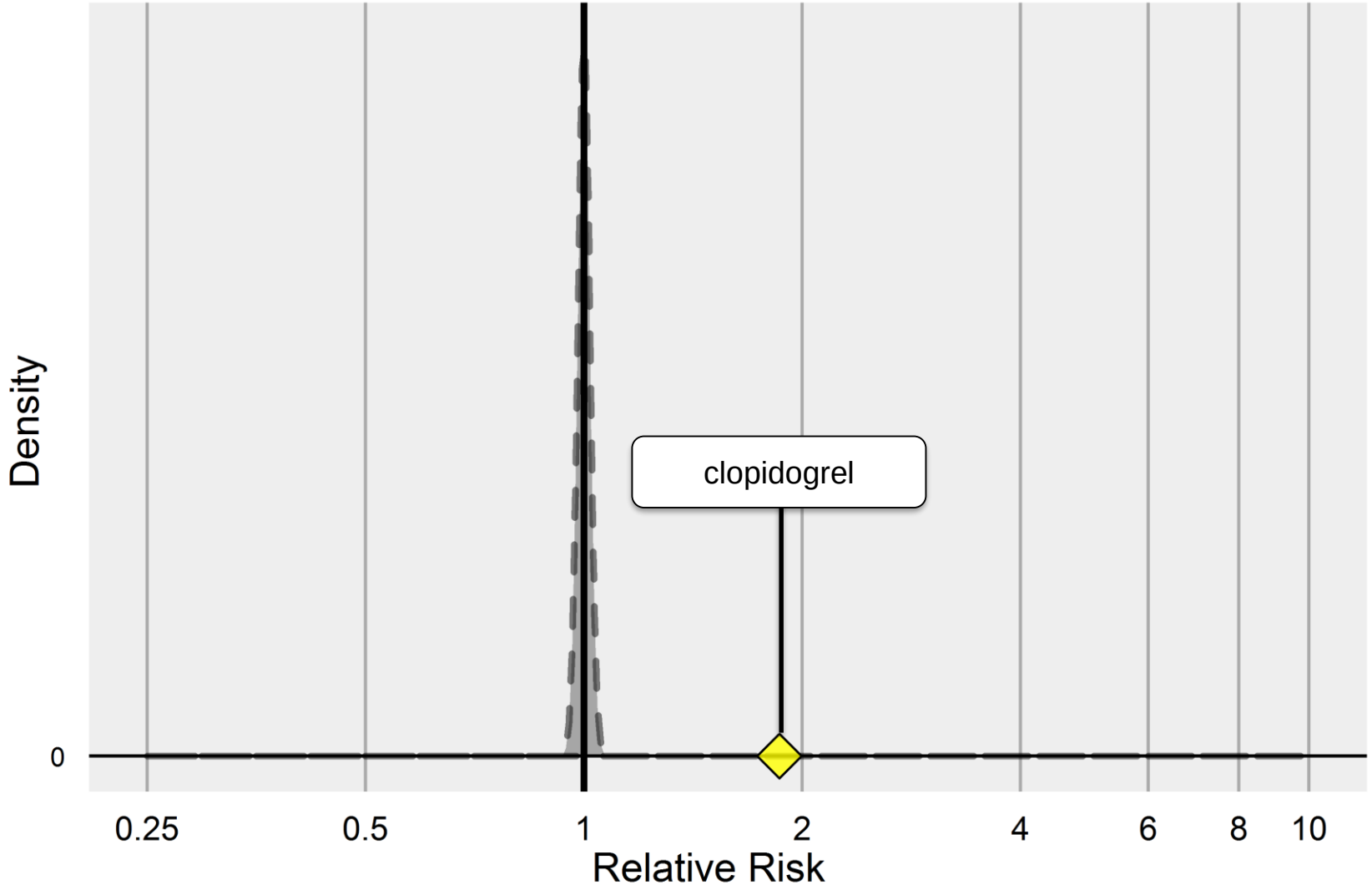
	Positive controls	Negative controls	Total
Acute Liver Injury	81	37	118
Acute Myocardial Infarction	36	66	102
Acute Renal Failure	24	64	88
Upper Gastrointestinal Bleeding	24	67	91
Total	165	234	399

阴性对照的标准:

- 事件没有被列入现存FDA结构化产品标签
- 在Tisdale et al，2010：药物诱导的疾病中没有被列为'致病因子’的药物
- 文献建设没有发现有力的研究证据证明相关性

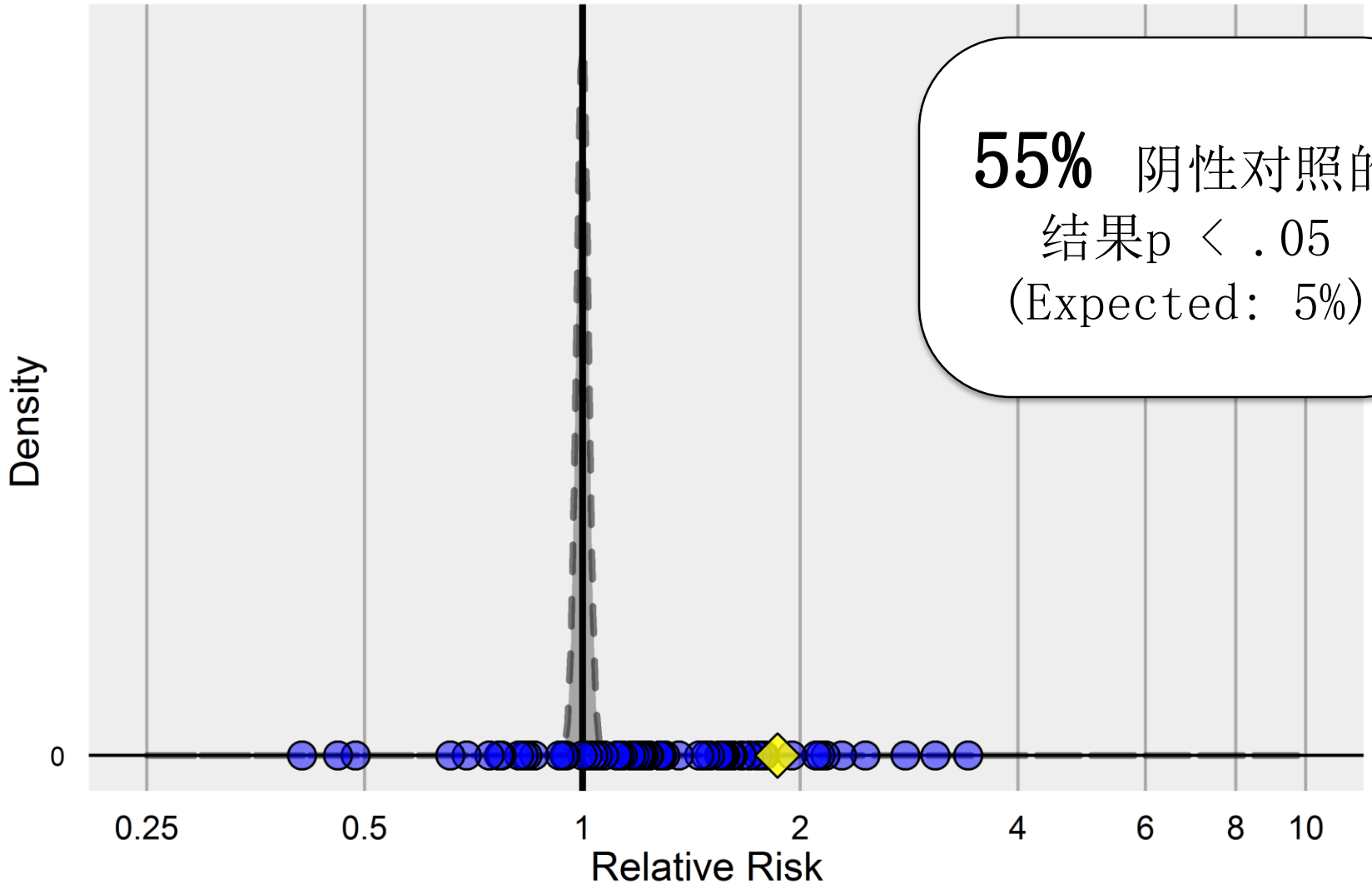
阴性对照& null 分布

CC: 2000314, CCAE, GI Bleed



阴性对照& null 分布

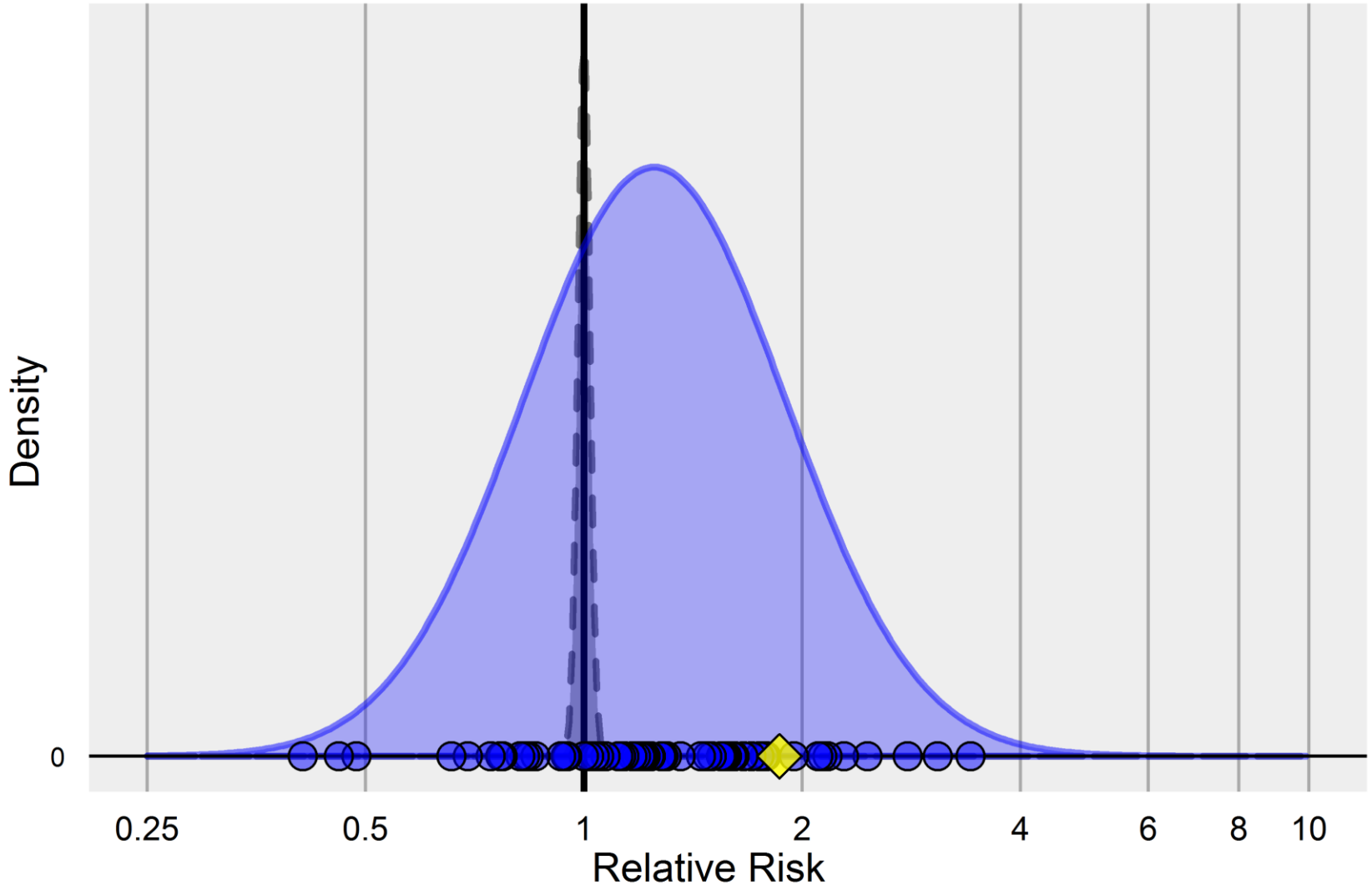
CC: 2000314, CCAE, GI Bleed



55% 阴性对照的结果p < .05 (Expected: 5%)

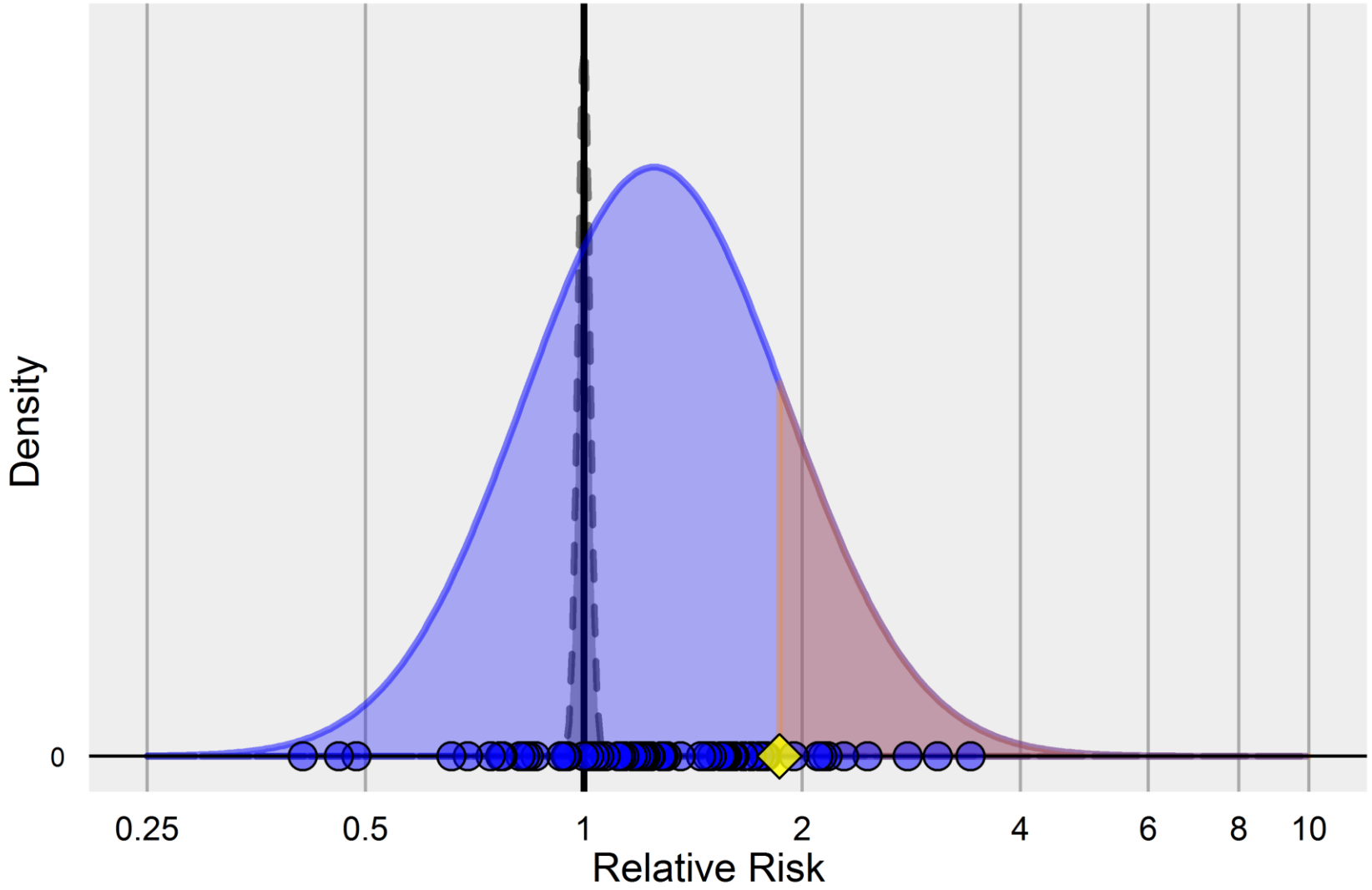
阴性对照& null 分布

CC: 2000314, CCAE, GI Bleed



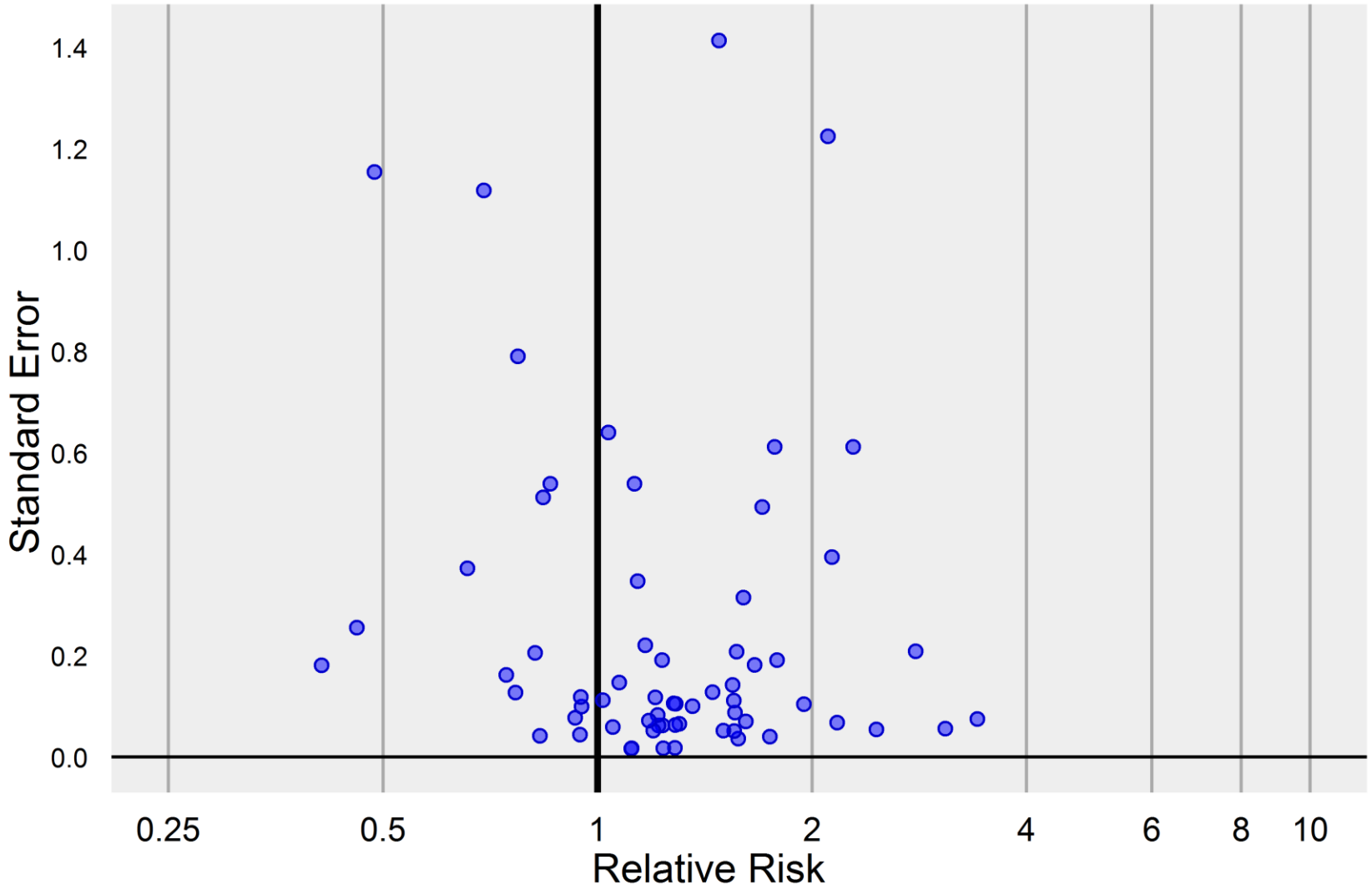
阴性对照& null 分布

CC: 2000314, CCAE, GI Bleed



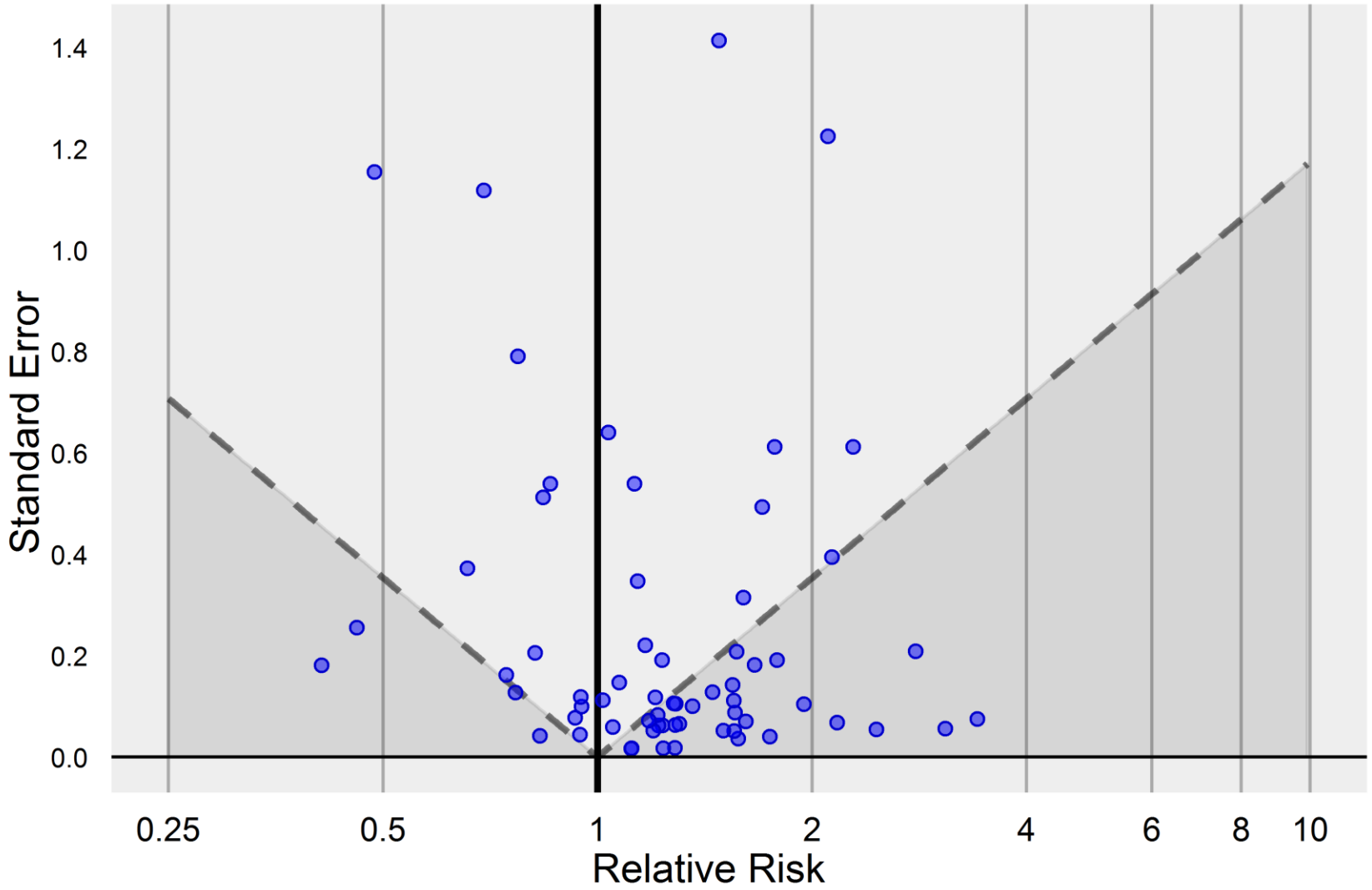
p-value 校正图

CC: 2000314, CCAE, GI Bleed



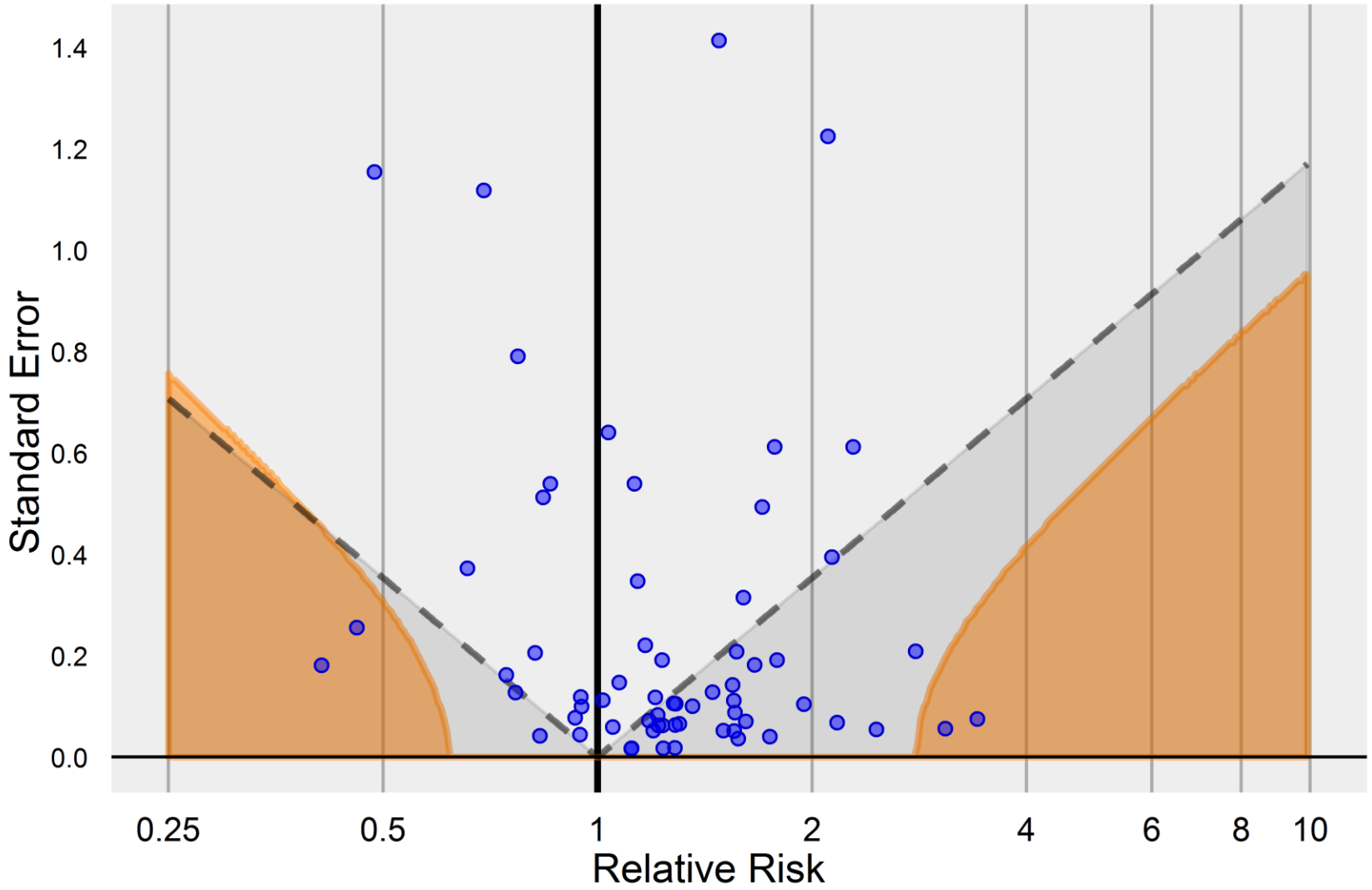
p-value 校正图

CC: 2000314, CCAE, GI Bleed



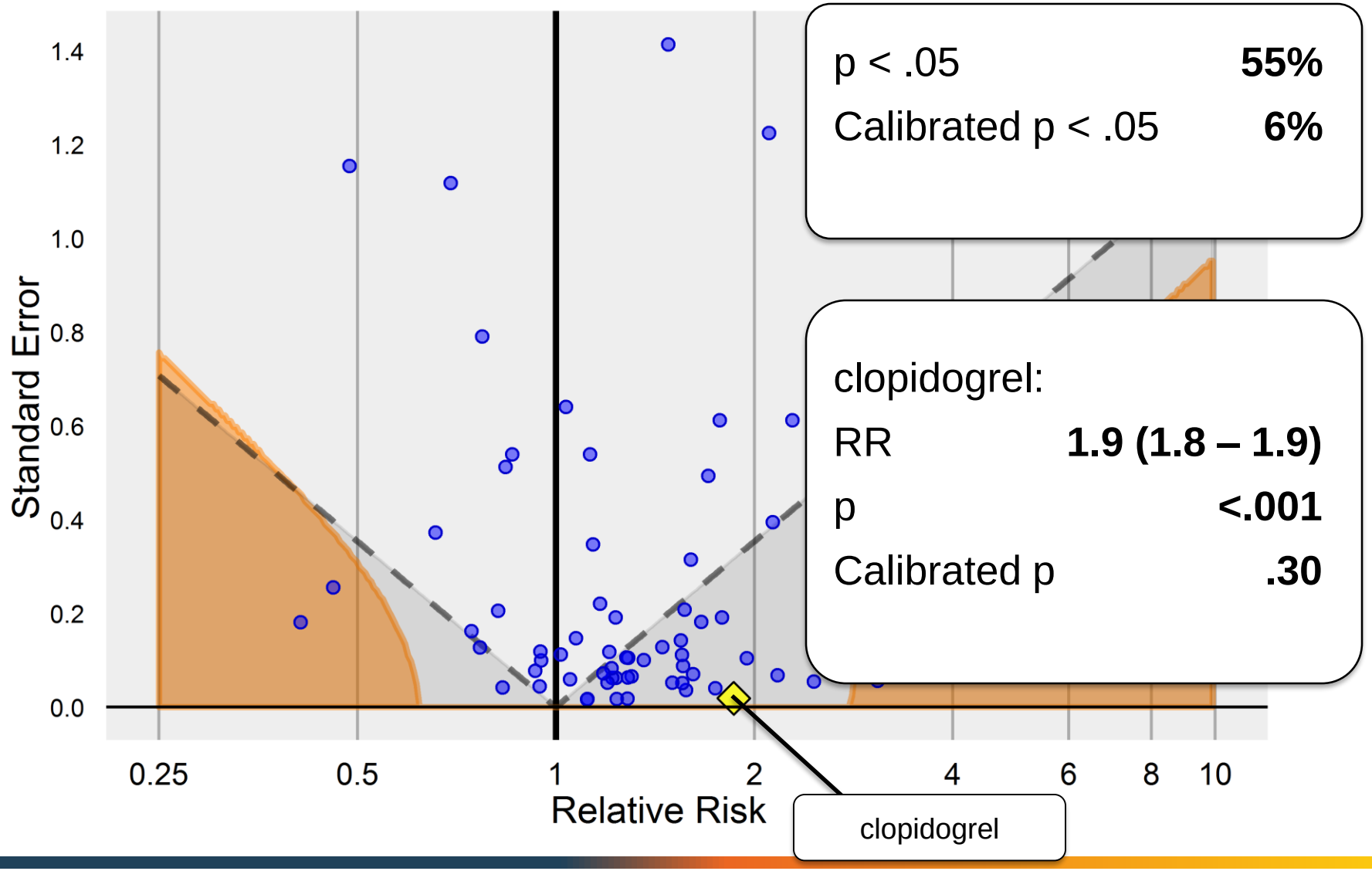
p-value 校正图

CC: 2000314, CCAE, GI Bleed



p-value 校正图

CC: 2000314, CCAE, GI Bleed



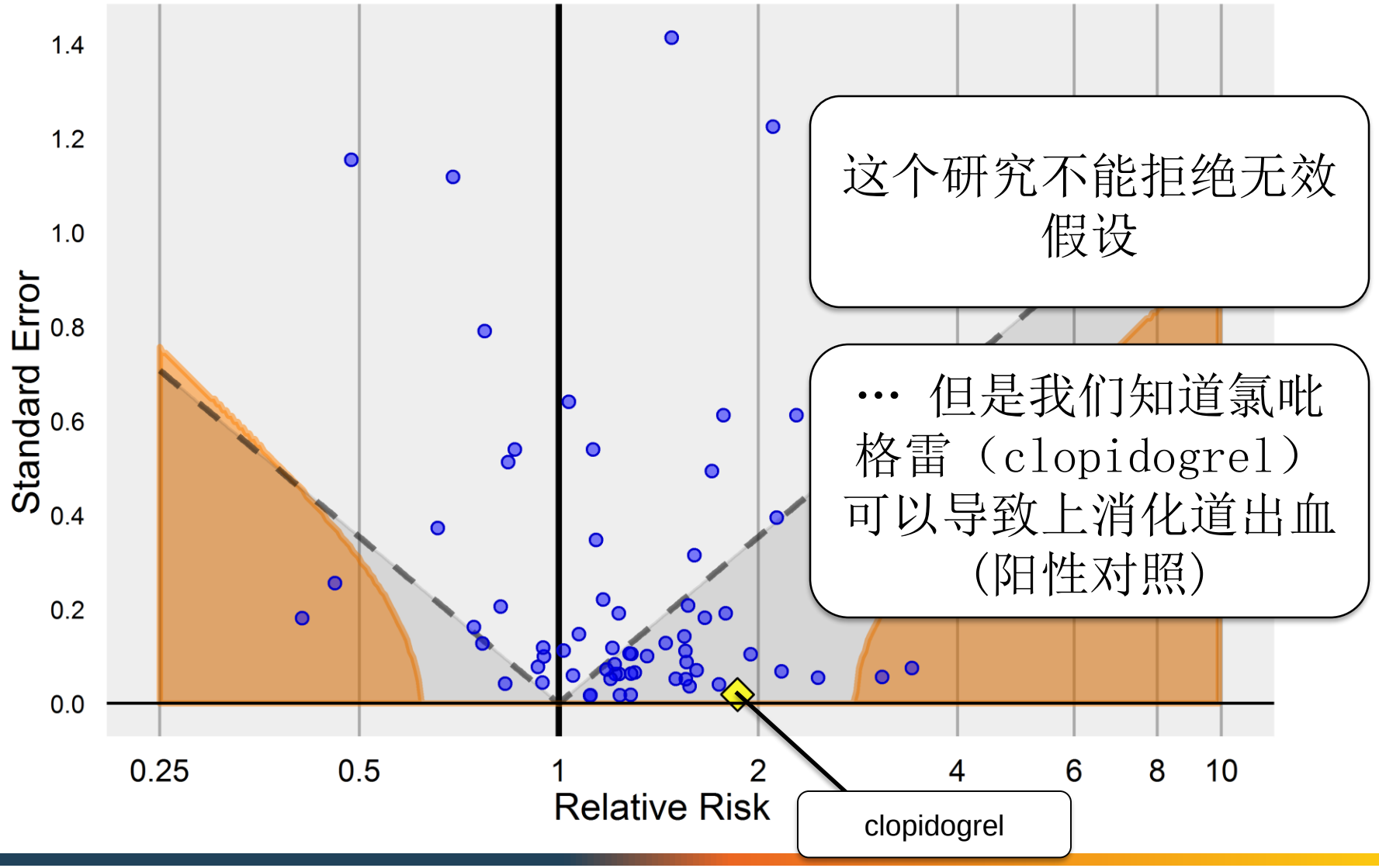
p < .05	**55%**
Calibrated p < .05	**6%**

clopidogrel:

RR	**1.9 (1.8 – 1.9)**
p	**<.001**
Calibrated p	**.30**

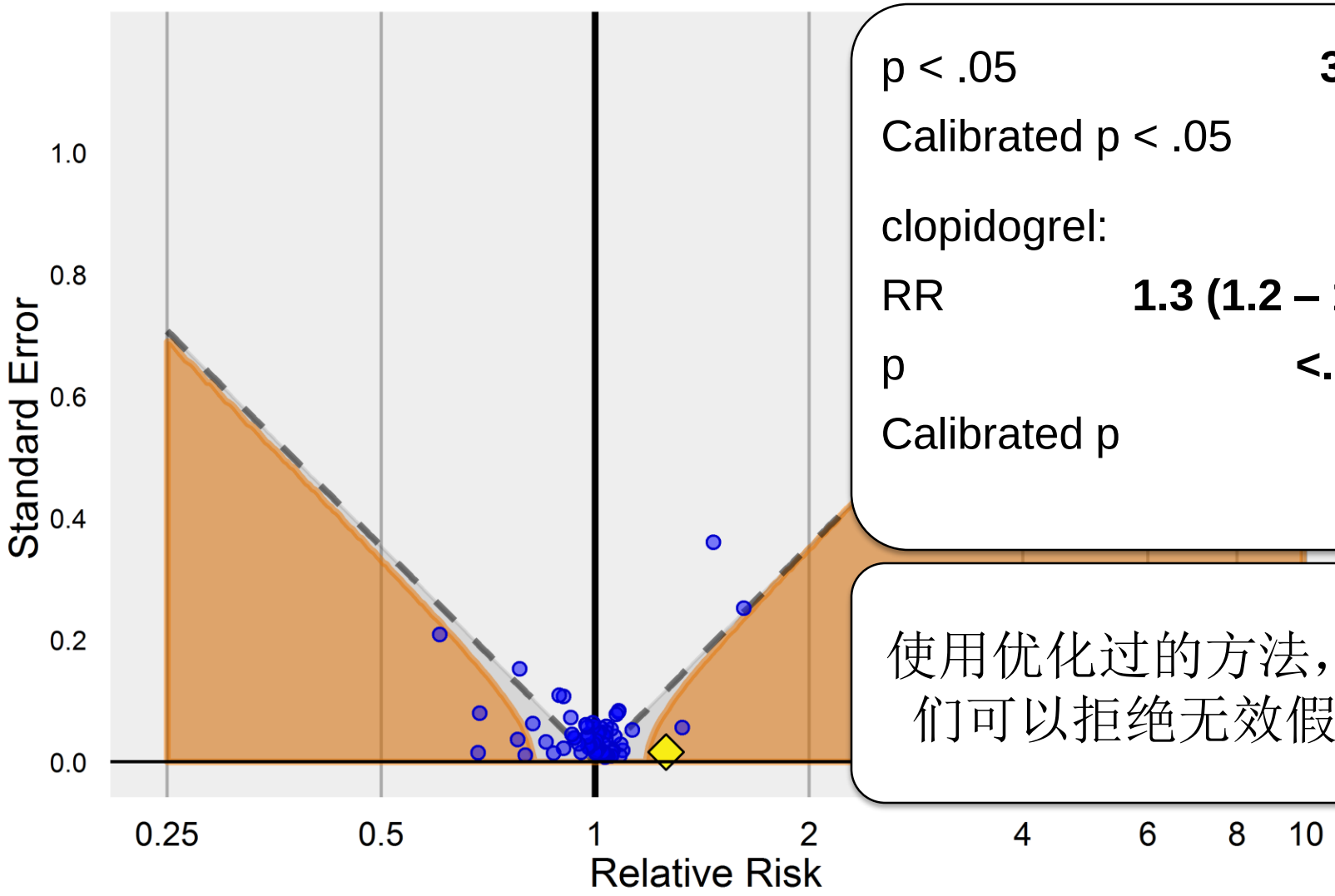
p-value 校正图

CC: 2000314, CCAE, GI Bleed



p-value 校正图

Optimal method: SCCS:1931010, CCAE, GI Bleed

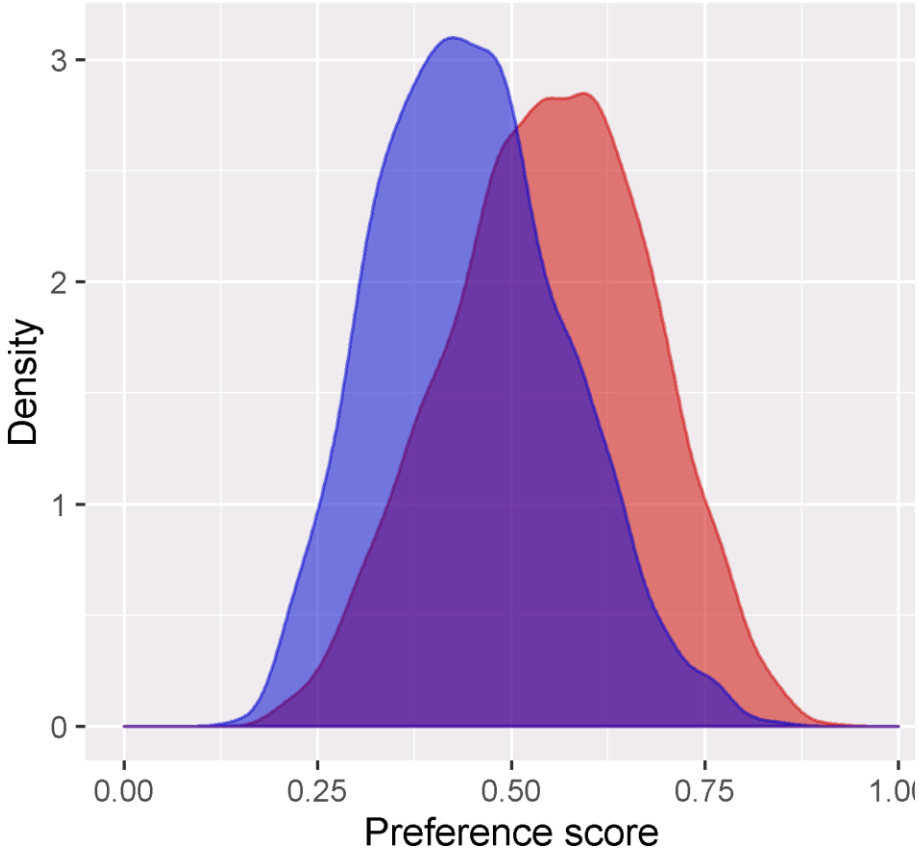


p < .05	**33%**
Calibrated p < .05	**9%**
clopidogrel:	
RR	**1.3 (1.2 – 1.3)**
p	**<.001**
Calibrated p	**.01**

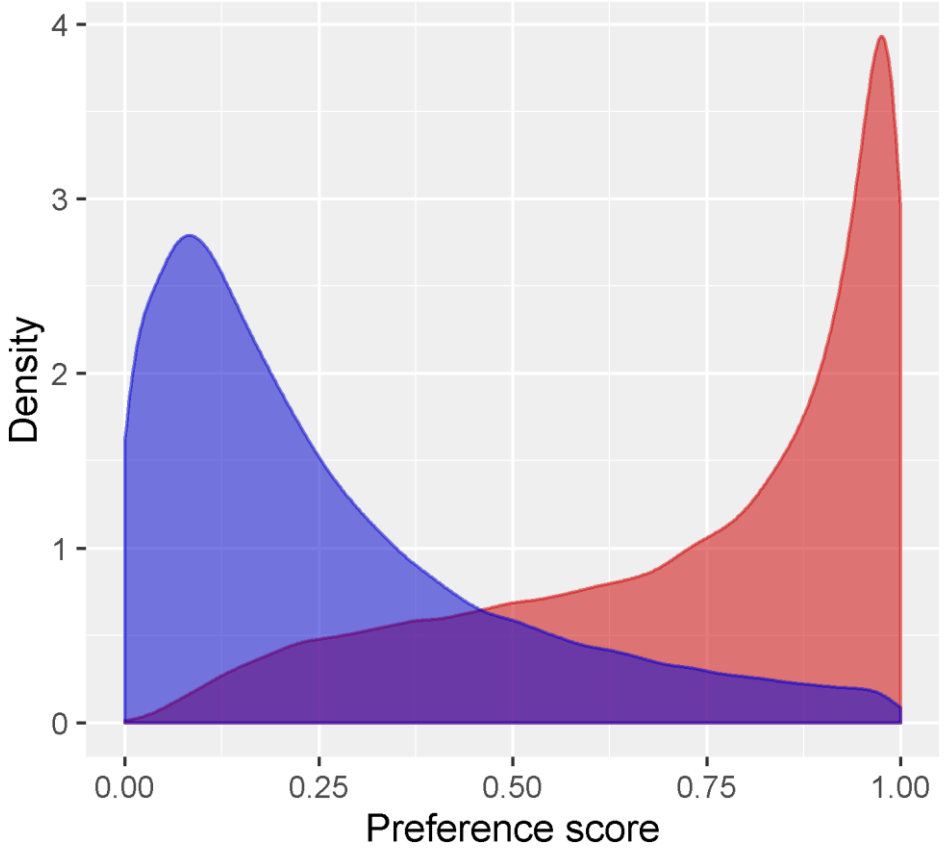
使用优化过的方法，我们可以拒绝无效假设

诊断：Propensity Score 校正

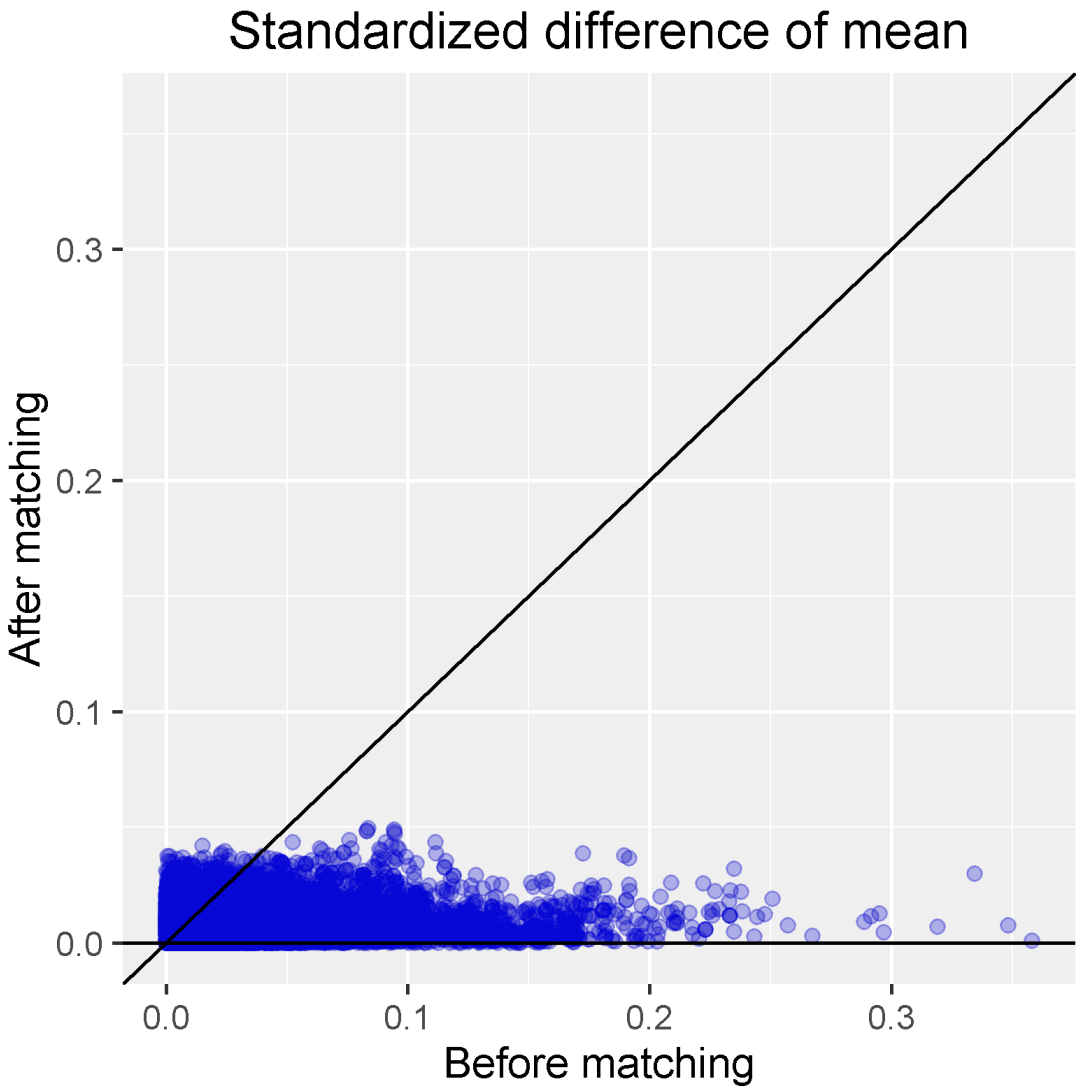
Good:



Bad:

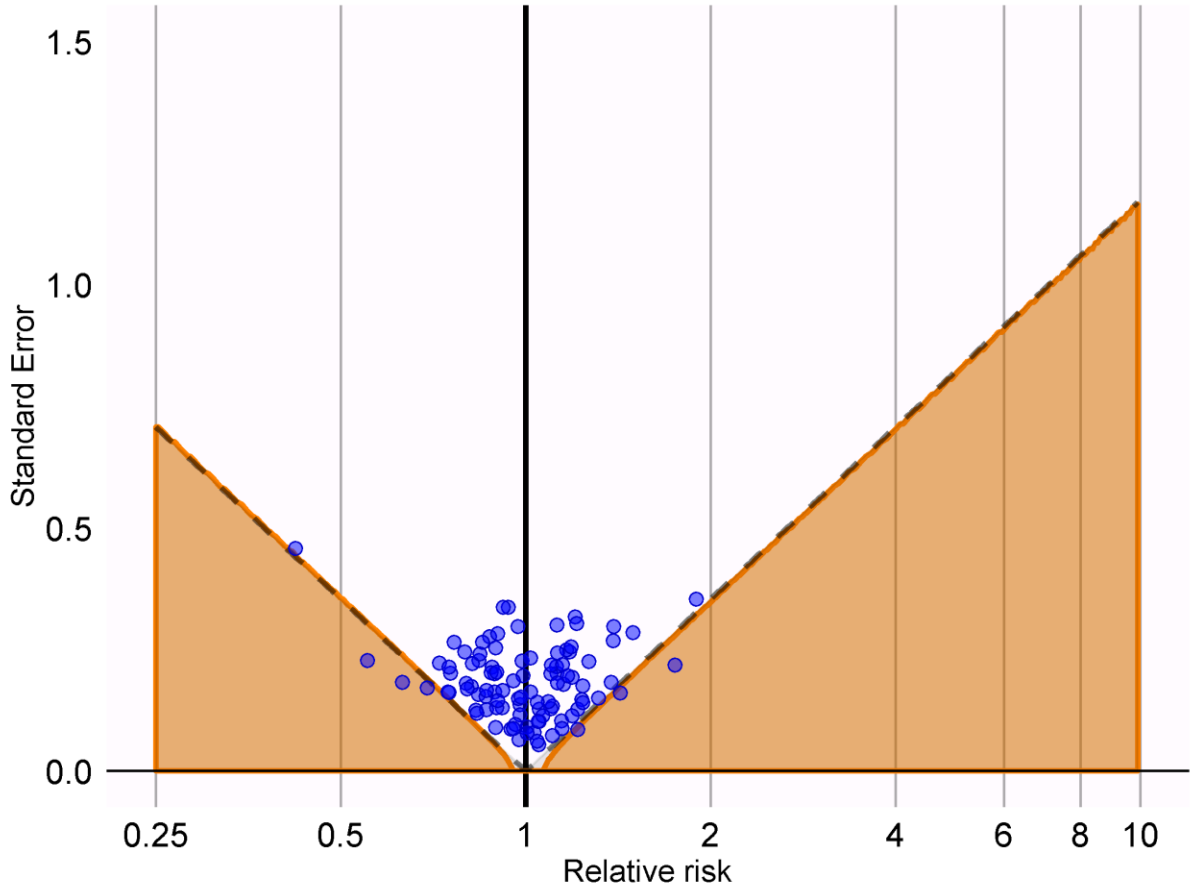


诊断：标准均数差


Standardized difference of mean
After matching
Before matching
0.0
0.1
0.2
0.3
0.0
0.1
0.2
0.3

诊断：阴性对照分布

约有95%的估计应落入95%置信区间



适合打印的报告格式

Population Level Effect Estimation

OHDSI estimation tutorial: Garbe replication: celecoxib vs. diclofenac for rate of upper gastro | Save | Close

Specification | Export

Print Friendly | R Code

Research question

To compare the risk of **OHDSI estimation tutorial: Garbe replication: outcome cohort - Upper gastrointestinal complication (UGIC) events** between **OHDSI estimation tutorial: Garbe replication: target cohort - celecoxib new users** and **OHDSI estimation tutorial: Garbe replication: comparator cohort - diclofenac new users**, we will estimate the population-level effect of exposure on the rate of the outcome during the period from 0 days from cohort start date to 0 days from cohort end date.

Study Design:

This study will follow a retrospective, observational, comparative cohort design. We define 'retrospective' to mean the study will be conducted using data already collected prior to the start of the study. We define 'observational' to mean there is no intervention or treatment assignment imposed by the study. We define 'cohort' to mean a set of patients satisfying a one or more inclusion criteria for a duration of time. We define 'comparative cohort design' to mean the formal comparison between two cohorts, a target cohort and comparator cohort, for the risk of an outcome during a defined time period after cohort entry.

In this study, we compare **OHDSI estimation tutorial: Garbe replication: target cohort - celecoxib new users** with **OHDSI estimation tutorial: Garbe replication: comparator cohort - diclofenac new users** for the rate of **OHDSI estimation tutorial: Garbe replication: outcome cohort - Upper gastrointestinal complication (UGIC) events** from 0 days from cohort start date to 0 days from cohort end date.

The overall study population could be considered to be patients who entered either the target cohort or comparator cohort. Patients were excluded from consideration is they qualified for both the target cohort and comparator cohort at any time in their record.

The rate of outcomes among patients in the target and comparator cohorts is determined by counting the number of outcome occurrences of **OHDSI estimation tutorial: Garbe replication: outcome cohort - Upper gastrointestinal complication (UGIC) events** during the time-at-risk of 0 days from cohort start date to 0 days from cohort end date.

Propensity scores will be used as an analytic strategy to reduce potential confounding due to imbalance between the target and comparator cohorts in baseline covariates. The propensity score is the probability of a patient being classified in the target cohort vs. the comparator cohort, given a set of observed covariates. In this study, the propensity score is estimated for each patient, using the predicted probability from a regularized logistic regression model, fit with a Laplace prior (LASSO) and the regularization hyperparameter selected by optimizing the likelihood in a 10-fold cross validation, using a starting variance of 0.01 and a tolerance of 2e-7.

The types of baseline covariates used to fit the propensity score model will be:

- Demographics
  - Gender
  - Age group (5-year bands)
  - Index year
- Conditions
  - In prior 365d

练习

- 为Graham的文章创建人群估计（population estimation）