– – OHDSI

Observational Health Data Sciences and Informatics (OHDSI)

2025-03-07

Contents

□□ **ix**
□□□□□□ ix
□□□□□ ix
□□□ ix
□□□□□□□□□□□□ x
□□□□□ x
□□□□□□□□□□ x

I OHDSI □□□□□□ 1

1 –□□□□□– OHDSI □□□□□□ 3
1.1 □□□□□□□□□□□□□ 3
1.2 □□□□□□□□□□□□□□□□□□□OMOP□ 3
1.3 □□□□□□□□□□□□□□□□□□□OHDSI 4
1.4 OHDSI□□□ 5
1.5 OHDSI□□□□□□ 6
1.6 □□□ 6

2 –□□□□□– □□□□□□□□□□ 7
2.1 □□□□□□□ 7
2.2 □□□□□□□□□□ 11
2.3 □□□ 11

3 –□□□□□– □□□□□□□□□□ 13
3.1 □□□□□□□□□ 13
3.2 □□□□□□□□□□□□: the Study-a-Thon 14
3.3 □□□□□□□□□□ 14
3.4 □□□□□□□ 14
3.5 □□□□□□□ 14
3.6 □□□□□□□□□□ 14
3.7 OHDSI□FAIR□□□□□□□□□□□□□ 15

II –□□□□□– □□□□□□□□□ 17

4 □□□□□□□□□ 19
4.1 □□□□□□□□ 19
4.2 □□□□□□□□□□ 21
4.3 CDM□□□□□□□□ 24
4.4 □□□□ 31
4.5 □□□ 31
4.6 □□ 31

5 □5□ –□□□□□– □□□□□□□□□□ 33
5.1 □□□□□□□□□□□□□□□□□□□□□□□□□□□□ 33
5.2 □□□□□ 35
5.3 □□ 39
5.4 □□ 41
5.5 □□□□□□□□□ 43
5.6 □□□□□ 43
5.7 □□□ 44
5.8 □□ 44

6 –□□□□□– □□□□□□□□□□ 45
6.1 □□□□ 45
6.2 □□□□ 1: ETL□□□□□ 45
6.3 □□□□2: □□□□□□□□□□□□ 52
6.4 □□□□ 3: ETL□□□ 57
6.5 □□□□ 4: □□□□ 57
6.6 ETL□□□□THEMIS 58
6.7 CDM□□□ETL□□□□□□□□ 58
6.8 ETL□□□□□□□□□□□ 58
6.9 □□□ 59
6.10 □□ 59

III –□□□□□– □□□□□ 61

7 □□□□□□□□□□ 63
7.1 □□□□ 63
7.2 □□□□□□□□□ 64
7.3 □□□□□□□□□ 65
7.4 □□□□□□□□□□□□□□□□ 65
7.5 □□□□□□□□ 66
7.6 □□□ 66
7.7 □□ 67

8 –□□□□□– OHDSI □□□□□ 69

8.1 □□□□□ 69
8.2 □□□□ 70
8.3 ATLAS 70
8.4 Methods Library 72
8.5 □□□□ 77
8.6 □□□ 78

9 –□□□□□– SQL□R 81
9.1 SqlRender 82
9.2 DatabaseConnector 88
9.3 CDM□□□□□ 91
9.4 □□□□□□□□□□□□□□□□□□□ 94
9.5 QueryLibrary 94
9.6 □□□□□□□□□□ 95
9.7 SQL□R□□□□□□□□□□ 96
9.8 □□□ 101
9.9 □□ 101

10 –□□□□□– □□□□□□□ 103
10.1 □□□□□□□ 103
10.2 □□□□□□□□□□□□□ 104
10.3 □□□□□□□□ 105
10.4 □□□□□□□□□ 105
10.5 □□□□□□□□□□ 106
10.6 □□□□□□□□□□ 106
10.7 ATLAS□□□□□□□□□□□ 106
10.8 SQL□□□□□□□□□□□□ 114
10.9 □□ 121
10.10□□ 121

11 –□□□□□– □□□□ 123
11.1 □□□□□□□□□□□□□□ 123
11.2 □□□□□□□□ 123
11.3 □□□□ 124
11.4 □□□ 124
11.5 □□□□□□□□□□ 125
11.6 ATLAS□□□□□□□□□□□□□□□ 125
11.7 ATLAS□□□□□□□□□□□□ 125
11.8 R□□□□□□□□□□□□ 134
11.9 ATLAS□□□□□□□□□□□□ 136
11.10ATLAS □□□□□□□□□□ 140
11.11□□□ 143
11.12□□ 143

12 □□□□□□□□ 145

12.1 □□□□□□□□□□ 145
12.2 □□□□□□□□□□□□ 147
12.3 □□□□□□□□ 147
12.4 □□□□□□□□□□□□□□□ 148
12.5 □□□□□□□□□□□□□□ 149
12.6 □□□□□□□□□□ 150
12.7 ATLAS□□□□□□□□□□ 151
12.8 R□□□□□□□□□□ 160
12.9 □□□□□ 166
12.10□□□ 170
12.11□□ 171

13 □□□□□□□ 173
13.1 □□□□ 173
13.2 □□□□□ 174
13.3 □□□□□□ 175
13.4 □□□□□□□□ 178
13.5 □□□□□□□□□□□□ 180
13.6 ATLAS□□□□□□□ 182
13.7 R□□□□□□ 193
13.8 □□□□□□□ 198
13.9 □□□□□□□□□□□□ 207
13.10□□□ 207
13.11□□ 207

(PART) □□□□□□□ 209

14 □□□□□□□ 211
14.1 □□□□□□□□□□□□□ 211
14.2 □□□□□□□□□□ 212
14.3 □□□□□□□□□□ 212

15 –□□□□□– □□□□□ 213
15.1 □□□□□□□□□□ 213
15.2 □□□□□□□□□ 214
15.3 □□□□□□□□□ 217
15.4 □□□□□□ACHILLES 218
15.5 Data Quality Dashboard□□□ 219
15.6 □□□□□□□□□□□□□ 220
15.7 □□□ 222
15.8 □□ 222

16 –□□□□□– □□□□□□ 225
16.1 □□□□□□□□□□□□ 225
16.2 □□□□□□□□□□□□ 225

16.3 □□□□□□□□□ 226
16.4 PheValuator 227
16.5 □□□□□□□□□□□□ 234
16.6 □□□ 234

17 –□□□□□– □□□□□□□□□□ 235
17.1 □□□□□□□□□□ 235
17.2 Methods Librar□□□□□□□□□□□□□ 236
17.3 Methods Library□□□□ 237
17.4 □□□ 238

18 –□□□□□– □□□□□□ 239
18.1 □□□□□□□□□□ 239
18.2 □□□□□□□□□ 240
18.3 □□□□□□□□□□□□ 244
18.4 OHDSI□□□□□□□□□□□□□ 249
18.5 □□□ 249

IV OHDSI□□ 251

19 –□□□□□– □□□□□ 253
19.1 □□□□□□□□□□□□□□□□□□□□ 253
19.2 □□□□□□□□ 254
19.3 □□□ 256

20 –□□□□□– OHDSI □□□□□□□□ 257
20.1 OHDSI □□□□□□□□□□□□ 257
20.2 OHDSI□□□□□□□□□ 257
20.3 OHDSI□□□□□□□□□□□□ 259
20.4 □□: □□□□□□□□□□□□□□□□□□ 260
20.5 OHDSI□□□□□□□□□□□□□□□□□□□□ 261
20.6 □□□ 261

Appendix 262

A –□□□□□– □□□ 263

B □□□□□□ 267
B.1 ACE□□□ 267
B.2 ACE□□□□□□□□□□□□□□□ 268
B.3 □□□□□□□AMI□ 271
B.4 □□□□□ 272
B.5 □□□□□□□□□□□□□□□□□□□□□□□□□□ 273
B.6 □□□□□□□□□□□□□□□□□□□□□□□ 275
B.7 □□□□□3□□□□□□□□□□□□□□□□□□□□□□□□□□□□ 278

B.8 ACE□□□□□□ 279
B.9 □□□□□□□□□□□□□□□□ARB□□□□ 279
B.10 □□□□□□□□□□□□□□□□□□□□□□□□ 281
B.11 □□□□□□□□□□□□□□□□□□□□□□□□DCCB□□□□ 281
B.12 □□□□□□□□□□□□□□□□□□□□□□□□□□NDCCB□□□□ 282
B.13 β□□□□□ 282
B.14 □□□□□□□□ 283
B.15 □□□□□□□□□□□□□ 283
B.16 α1□□□□□ 283

C □□□□ 285
C.1 ACE□□□□□□□□□□□□□□□□□□□□□□□□ 285

D □□□□□□□□□□□□ 289

E □□□ 291
E.1 □□□□□□□□ 291
E.2 □□□□□□□□□□□□□ 294
E.3 ETL (Extract-Transform-Load) 295
E.4 □□□□□□□□□□□□□ 296
E.5 SQL□R 296
E.6 □□□□□□□ 298
E.7 □□□□□ 303
E.8 □□□□□□□□ 310
E.9 □□□□□□□□ 314
E.10 □□□□□ 316
E.11 317

Bibliography 319

Index 327

□□□□OHDSI □□□□□□□□□□□□□□□□□□□□□□□□□□OHDSI□□□□□□□□□□□□□□□□OHD
http://book.ohdsi.org □□□□□□□□□□□□□□□□□□□□

Hamed Abedtash	Mustafa Ascha	Mark Beno
Clair Blacketer	David Blatt	Brian Christian
Gino Cloft	Frank DeFalco	Sara Dempster
Jon Duke	Sergio Eslava	Clark Evans
Thomas Falconer	George Hripcsak	Vojtech Huser
Mark Khayter	Greg Klebanov	Kristin Kostka
Bob Lanese	Wanda Lattimore	Chun Li
David Madigan	Sindhoosha Malay	Harry Menegay
Akihiko Nishimura	Ellen Palmer	Nirav Patil
Jose Posada	Nicole Pratt	Dani Prieto-Alhambra
Christian Reich	Jenna Reps	Peter Rijnbeek
Patrick Ryan	Craig Sachson	Izzy Saridakis
Paola Saroufim	Martijn Schuemie	Sarah Seager
Anthony Sena	Sunah Song	Matt Spotnitz
Marc Suchard	Joel Swerdel	Devin Tian
Don Torok	Kees van Bochove	Mui Van Zandt
Erica Voss	Kristin Waite	Mike Warfe
Jamie Weaver	James Wiggins	Andrew Williams
Seng Chan You		

□□□□□□□□OHDSI□□□□□□□□

Table 1: □□□□□□□□□□□Methods Library□□□□□□□□□□□□□

□□□□□	□□□□□
CaseControl	1.6.0
CaseCrossover	1.1.0
CohortMethod	3.1.0
Cyclops	2.0.2
DatabaseConnector	2.4.1
EmpiricalCalibration	2.0.0
EvidenceSynthesis	0.0.4
FeatureExtraction	2.2.4
MethodEvaluation	1.1.0
ParallelLogger	1.1.0
PatientLevelPrediction	3.0.6
SelfControlledCaseSeries	1.4.0
SelfControlledCohort	1.5.0
SqlRender	1.6.2

Part I

OHDSI

Chapter 1

– – OHDSI

□□□□□*Patrick Ryan & George Hripcsak*

□□□□□

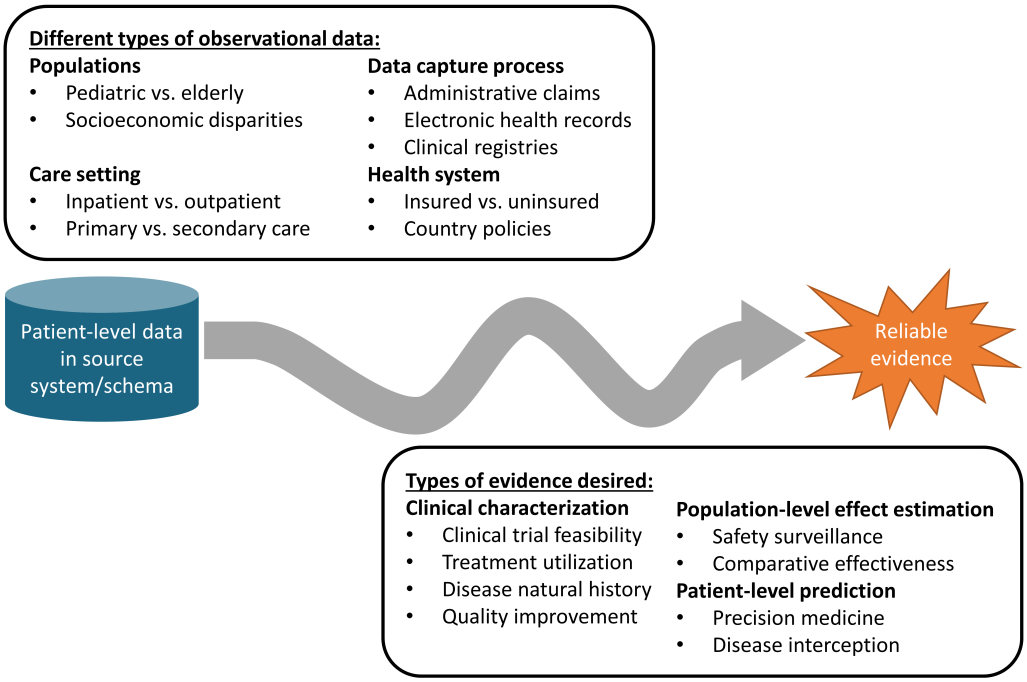


Figure 1.1: □□□□□□□□□□□□□□□

OMOP□□□□□□□□□□□□□□□(Ryan et al., 2012, 2013b) □□□□□□□□(Schuemie et al., 2014)□□□□□□□□□□□□□□□

1.3.2

□□□

OHDSI□□□□□□□□□□□□□OMOP CDM□□□□□□□□□□□□□

Chapter 2

— —

□□□□□*Hamed Abedtash & Kristin Kostka*

□□□□□□□□□□□□ - □□

OHDSI□□□□□□□□□□□□

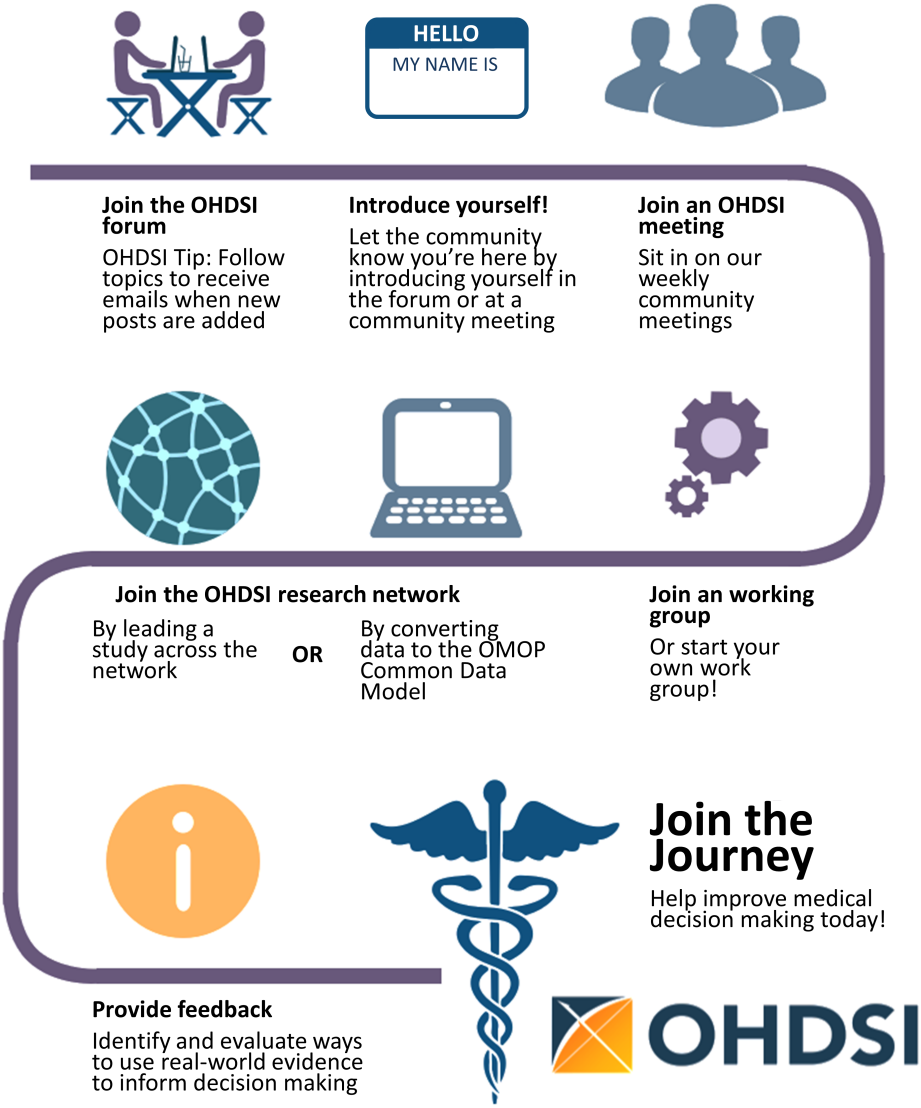


Figure 2.1: □□□□□□□□ — OHDSI□□□□□□□□□□□□□□□□

□□□□□□□

□□□□□□□□□	□□	□□□□□
FHIR □□□□□□□□	OHDSI FHIR□□□□□□□□□□□□□□□□OHDSI□□□□□□□□□□□□□□EHR□□□	□□□□□□□□□□□□□□□□□□□□□□□□□□
GIS	OMOP CDM□□□□□□□□□□□□□□□□□□□

2.2

□□□□□□□□□□□□□□□□□OHDSI □□□□□□□□□□□□□□□□□□

Chapter 3

— —

□□□□□*Kees van Bochove*

OHDSI□□□□□

3.2 : the Study-a-Thon

□□□□□□□□□□□□□□□□□□□□□□study-a-thon□□□□□□□□□□□□□□□□□□□□OMOP□□□□□□□□C
a-thon□□□□□□□□□□□□□□□□□□□□□□2□□□□□□□□□□□□□□□□□□□

3.7 OHDSI FAIR

3.7.1

□□□□□□□□□□□□□□Wilkinson et al. (2016) □□□□□□□FAIR□□□□OHDSI□□□□□□□□□□□□□□

3.7.2

OMOP□□□□□□□□□□□□□□□□

Part II

Chapter 4

□□□□: *Clair Blacketer*

□□□□

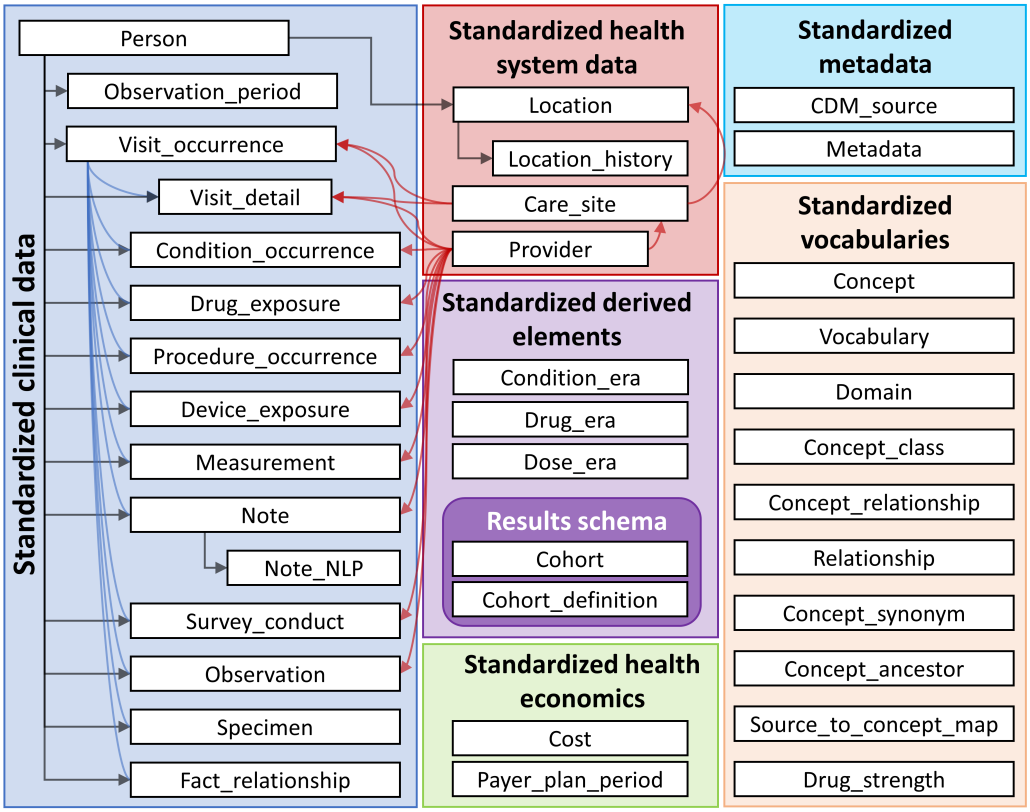


Figure 4.1: CDM□□□□□6.0□□□□□□□□

- □□□□□□□□: □□□□□□□□□□□

□□□□□□	□□□□ID	□□□□□□	□□□□ID
257000	Procedure	158	Payer
163807	Condition	123	Visit
145898	Observation	51	Cost
89645	Measurement	50	Race
33759	Spec Anatomic Site	13	Plan Stop Reason
17302	Meas Value	11	Plan
1799	Specimen	6	Episode
1215	Provider Specialty	6	Sponsor
1046	Unit	5	Meas Value Operator
944	Metadata	3	Spec Disease Status
538	Revenue Code	2	Gender
336	Type Concept	2	Ethnicity
194	Relationship	1	Observation Type

4.2.5

CDM□□□□□□□

- □□□□□□□□□□□

TERM CONNECTIONS (82)

RELATIONSHIP	RELATES TO	CONCEPT ID	VOCABULARY
ICD-9-CM to MedDRA (MSSO)	Pulmonary tuberculosis	36110777	MedDRA
Non-standard to Standard map (OMOP)	Pulmonary tuberculosis	253954	SNOMED
Subsumes	Other specified pulmonary tuberculosis	44830894	ICD9CM
	Other specified pulmonary tuberculosis, bacteriological or histological examination not done	44836741	ICD9CM
	Other specified pulmonary tuberculosis, bacteriological or histological examination unknown (at present)	44836742	ICD9CM
	Other specified pulmonary tuberculosis, tubercle bacilli found (in sputum) by microscopy	44821641	ICD9CM
	Other specified pulmonary tuberculosis, tubercle bacilli not found (in sputum) by microscopy, but found by bacterial culture	44833188	ICD9CM

Figure 4.3: □□□□SNOMED□□□

4.3 CDM

CDM□□16□□□□□□□□□□□□□□10□□□□□□□□□□□□□□2□□□□□□□□□□□□□□4□□□□□□□□□□□□□□□□□□□□□□□□□CDM Wiki□□□□□□□□□□□□□□ [1]□

□□□□□□□□□□□□□□□□□□□

4.3.2 PERSON

Lauren□□□□□□□□□□□□□

- □□□36□□□□□□□
- □□□□□□□□1982□3□12□□□
- □□□□□□□□
- □□□□□□□□□□□

□□□□□□□□□□□□□PERSON□□□□□□□□□□□□□□□□□□

Table 4.3: PERSON□□□□

□□	□	□□
PERSON_ID	1	PERSON_ID□□□□□□□□□□□□□□
GENDER_CONCEPT_ID	8532	□□□□□□□□□□□□□□□□ID□8532□
YEAR_OF_BIRTH	1982	
MONTH_OF_BIRTH	3	
DAY_OF_BIRTH	12	
BIRTH_DATETIME	1982-03-12 00:00:00	□□□□□□□□□□□□□□□□□□□□□□□□
DEATH_DATETIME		
RACE_CONCEPT_ID	8527	□□□□□□□□□□□ID□8527□□□□□□
ETHNICITY_CONCEPT_ID	38003564	□□□□□□□□□□□□□□□□□□□□

4.3.3 OBSERVATION_PERIOD

OBSERVATION_PERIOD□□□□□□□

Table 4.6: VISIT_OCCURRENCE□□□□□

□□	□	□□
VISIT_OCCURRENCE_ID	514	□□□□□□□□□

□□	□	□□
CONDITION_ OCCURRENCE_ID	964	□□□□□□□□

□□	□	□□
DRUG_CONCEPT_ID	1127433	□□□□□□□□□

□□	□	□□
DRUG_SOURCE_CONCEPT_ID	750264	□□□□□□□□□□□□□□□□□□□□□□□□□□□ 750264 ND-C□□□□”Acetaminophen 325 MG Oral Tablet□□□□□□□□□□□ 325 MG □□□□“□□□□□□□
ROUTE_SOURCE_VALUE	NULL	□□□□□□□□□□□□□□□□□□□□□□□□□□□

4.3.7 PROCEDURE_OCCURRENCE

PROCEDURE_OCCURRENCE□□□□□□□□□□□□□□□□□□□

Column name	Value	Explanation
VISIT_OCCURRENCE_ID	740	□□□□□□□□□□□□□□□□□□□□□□□□
VISIT_DETAIL_ID	NULL	□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□ VISIT_detail_id □□□□□□□□
PROCEDURE_SOURCE_VALUE	304435002	□□□□□□□□□□□□□□□□□□□□□□□□
PROCEDURE_SOURCE_CONCEPT_ID	4127451	□□□□□□□□□□□□□□□□□□□□□□□□
MODIFIER_SOURCE_VALUE	NULL	□□□□□□□□□□□□□□□□□□□□□□□□

4.4

□□□□□CDM□□□□□□□□□□□□□□□□□□□

□□□□

□□□3□□□□□□□□□□□□□□ 8.4.5 □□□□□□□□□□□□□R□R-Studio□□□□Java□□□□□□□□□□□□□

```
install.packages(c("SqlRender", "DatabaseConnector", "remotes"))
remotes::install_github("ohdsi/Eunomia", ref = "v1.0.0")
```

Eunomia□□□□□□□□□□□□□□R□□□□□□□□□□□□□□CDM□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

```
connectionDetails <- Eunomia::getEunomiaConnectionDetails()
```

CDM□□□□□□□□□□□□□□main□□□□□□□□CONDITION_OCCURRENCE□□□□□□□□□□□□□□□□□□□

Chapter 5

5 — —

□□□□: *Christian Reich & Anna Ostropolets*

OMOP□□□□

1660.

A General BILL for this present Year,

Ending the 11th Day of December 1660.

According to the Report made to the King's most excellent Majesty,

By the Company of Parish Clerks of LONDON, &c.

DISEASES and CASUALTIES.

Abortive and Stillborn	421	Flox and Small Pox	1523	Palsy	17
Aged	909	Found dead in the Streets, Fields, &c.	2	Plague	36
Ague and Fever	2303	French Pox	51	Plurisy	12
Apoplexy and Suddenly	91	Gout	4	Quinsy and sore Throat	21
Blasted and Planet	3	Grief	13	Rickets	441
Bleeding and bloody Issue	7	Griping in the Guts	253	Rising of the Lights	210
Bloody Flux, Scowring, and Flux	346	Hanged and made away themselves	11	Rupture	12
Burnt and Scalded	6	Head-ach and Headmouldshot	35	Scurvy	82
Cancer, Gangrene and Fistula	63	Jaundies	102	Shot	7
Canker, sore Mouth and Thrush	73	Imposthume	105	Shingles	1
Childbed	226	Killed by several Accidents	55	Sores, Ulcers, broken and bruised Limbs	61
Chrisomes and Infants	858	King's Evil	28	Spleen	7
Cold, Cough and Hiccough	33	Lethargy	6	Spotted Fever and Purples	368
Colick and Wind	116	Livergrown	8	Starved	7
Consumption and Tissick	2982	Lunatick and Frenzy	14	Strangury	22
Convulsion	742	Megrims	5	Stopping of the Stomach	186
Cut of the Stone and Stone	46	Measles	6	Surfeit	202
Dropsy and Tympany	646	Mother	1	Swine Pox	2
Drowned	57	Murthered	7	Teeth and Worms	839
Executed	7	Overlaid and Starved at Nurse	46	Vomiting	8
Falling Sickness	4			Wen	1

Figure 5.1: 1660□□□□□□□□□

5.1.2

□□□□□□□□□□□□□□□□□□□□□Pallas□□□□□□□□□□□□□□□□□□□□□□□ATHENA[4]□□□□□

OMOP CDM□□□zip□□□□□
5.2.6 □□□□□□□□□□□□□□□□□□□□

5.2.2

□□□□□□□□□1□□□□□□□□□

Table 5.1: □□□□□□

□□□□□□□□□

□□□□□ID	□□□□□□□□□□□□□□□	□□□□ID	□□□□□□□ID □□□□□□□□□

5.2.10

□□□□□□□□□□□□□□

Table 5.4: □□□□□□□□□□

□□ID□□	□□
“Maps to”□”Mapped from”	□□□□□□□□□□

□□□□□□□□□□□□□□□□OHDSI Wiki□□□□□□□□□□□□□[5]□

5.3.2

□□□□□□□Is a□-□Subsumes□□□□□□□□□□□□□□□□

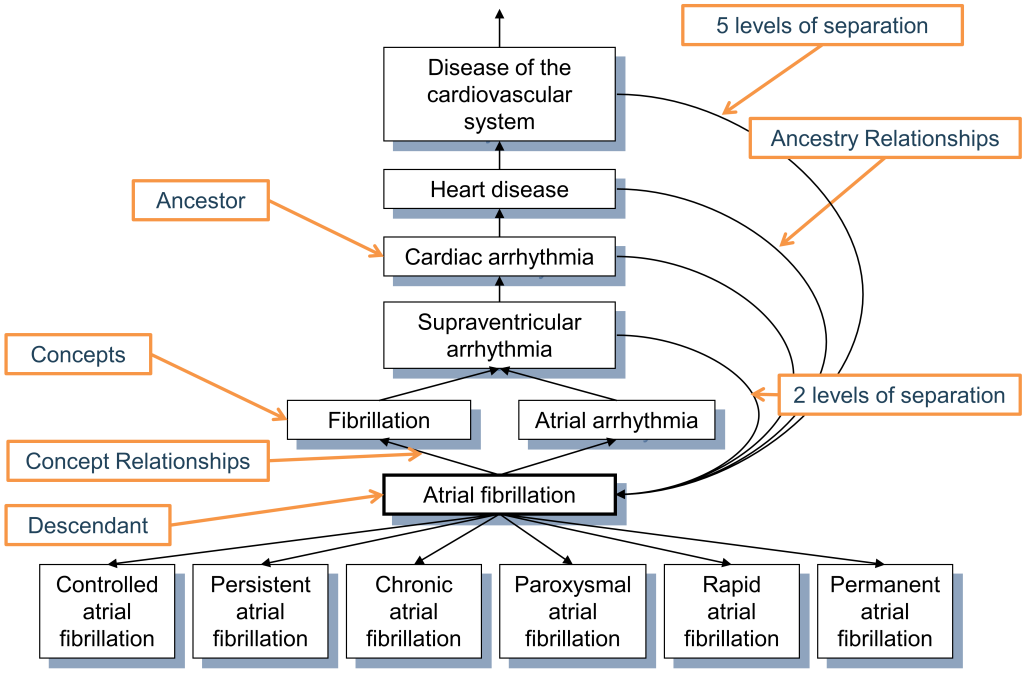


Figure 5.6: □□□□□□□□□□□□□□□□□□□□□□□□□□□□□Is a□□□Subsumes□□□CONCEPT_ANCESTOR□□

5.5

DOMAIN_ID□VOCABULARY_ID□CONCEPT_CLASS_ID□□□□CONCEPT□□□□□□□□□□
DOMAIN□VOCABULARY□CONCEPT_CLASS□RELATIONSHIP □□□□□□□□□□*_ID
□□□□□□□□□□□□□□□□□□□□□*_NAME □□□□□□□□CONCEPT □□□□□□□□□□□□□
*_CONCEPT_ID □□□□□□□□□□□□□□□CONCEPT □□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□VOCABULARY□□□VOCABUL

5.6

5.6.1

OMOP CDM□□□□□□□□□□□□□□□□□□□

5.6.9

□□□□□□□□□

Chapter 6

— —

□□□□: *Clair Blacketer & Erica Voss*

6.1

□□□□□/□□□□□□□OMOP□□□□□□□□□□CDM□□□□□□□□□□□□□□□□□□□□□□ETL□□□

ETL□□□4□□□□□□□□□□□□□□□

1. □□□□□□□□□CDM□□□□□□□□□ETL□□□□□□□□□
2. □□□□□□□□□□□□□□□□□□□□□□□□□□
3. □□□□ETL□□□□□□□
4. □□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□OHDSI□□□□□□□□□□□□□□□□□□□□□□□□□

6.2 1: ETL

ETL□□□□□□□□□□□□□□□□□□□□□□□□□□□□ETL□□□□□□□□□□□□□□□□CDM□□□□□□

ETL□□□□□□□□□□□□□□□□□□□□□□□□□□□□2□□□□□□□□□□□□□□□□White Rabbit□Rabbit-in-a-Hat□□□

6.2.1 White Rabbit

□□□□□□□□ETL□□□□□□□□□□□□□□□□□□□

□□□□□

White Rabbit□□□□□□□□□

□□□□□□□□□□□

White Rabbit□□□□□□□□□□□

	A	B	C	D	E	F	G
1	Table	Field	Type	Max length	N rows	N rows checked	Fraction empty
2	dbo.allergies	start	date	10	3184	3184	0
3	dbo.allergies	stop	date	10	3184	3184	0.725188442
4	dbo.allergies	patient	varchar	36	3184	3184	0
5	dbo.allergies	encounter	varchar	36	3184	3184	0
6	dbo.allergies	code	varchar	9	3184	3184	0
7	dbo.allergies	description	varchar	24	3184	3184	0
8							
9	dbo.careplans	id	varchar	36	30199	30199	0
10	dbo.careplans	start	date	10	30199	30199	0
11	dbo.careplans	stop	date	10	30199	30199	0.057849598
12	dbo.careplans	patient	varchar	36	30199	30199	0
13	dbo.careplans	encounter	varchar	36	30199	30199	0
14	dbo.careplans	code	varchar	15	30199	30199	0
15	dbo.careplans	description	varchar	62	30199	30199	0
16	dbo.careplans	reasoncode	varchar	9	30199	30199	0.050796384
17	dbo.careplans	reasondescription	varchar	56	30199	30199	0.050796384
18							

Overview | dbo.allergies | dbo.carepla ...

Figure 6.3: □□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□

□□□□□□

□□□□□□□□□□□□□□ETL□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

1. White Rabbit□□□□□□□□□□□□□□□□□
2. □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□CDM□□□□□□□□□□□□□
3. □□□□□□□□□□□□□CDM□□□□□□□□□□□□□□□□□□□□□□□□□□□□CDM□□□□□□
4. □□□□□□□□□□□CDM□□□□□□□□□□□□□□□□□□□□□CDM□□□□□□□□□□□□□□□□
5. Rabbit-In-a-Hat□□□□□□□□□□□MS Word□□□□□□□□□□□□□□□

ETL□□□□□□□□

White Rabbit□□□□□□□□□□Rabbit-In-a-Hat□□□□□□□□□□□□□OMOP CD-M□□□□□□□□□□□□□□□□□□□

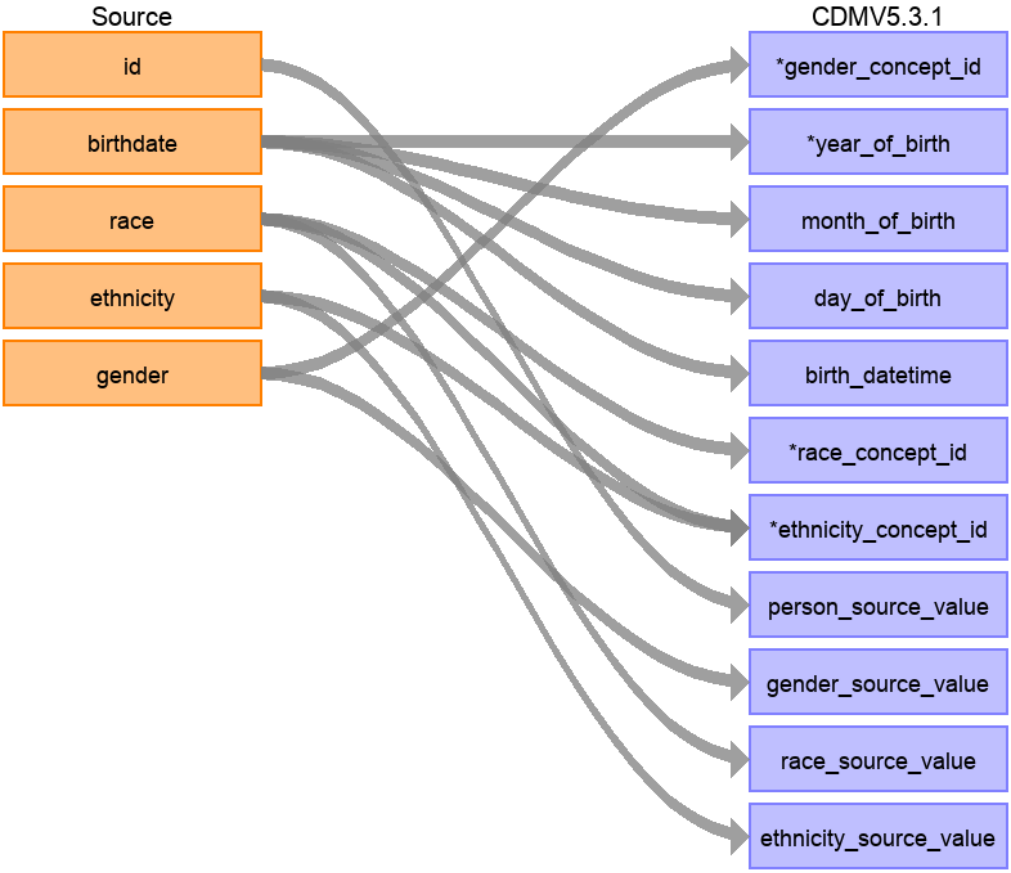


Figure 6.6: Synthea Patients□□□□□□CDM PERSON□□□□□□□□□□□□

□ 6.1 □□□Synthea patients□□□□□CDM PERSON□□□□□

□□□□□□□□	□□□□□□□□□	□□□□□□□□□□
GENDER_SOURCE_CONCEPT_ID		
RACE_SOURCE_VALUE	race	
RACE_SOURCE_CONCEPT_ID		
ETHNICITY_SOURCE_VALUE	ethnicity	□□□□□ETHNICITY_SOURCE_VALUE
ETHNICITY_SOURCE_CONCEPT_ID		

Synthea□□□□□□□CDM□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□[3]□

6.3 2:

OMOP□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□CDM□□□□□□□□□□□□□□□□□□□

6.3.1 Usagi

Usagi□□□□□□□□□□□

□□□□□Filters□□□□□□□□□□□□□□□Usagi□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□6
standard concepts□□□□□□□□□□□□□□□□□□□

“Include source terms” (□□□□□□□□□□) □□□□□□□□□□□Usagi □□□□□□□□□□□□

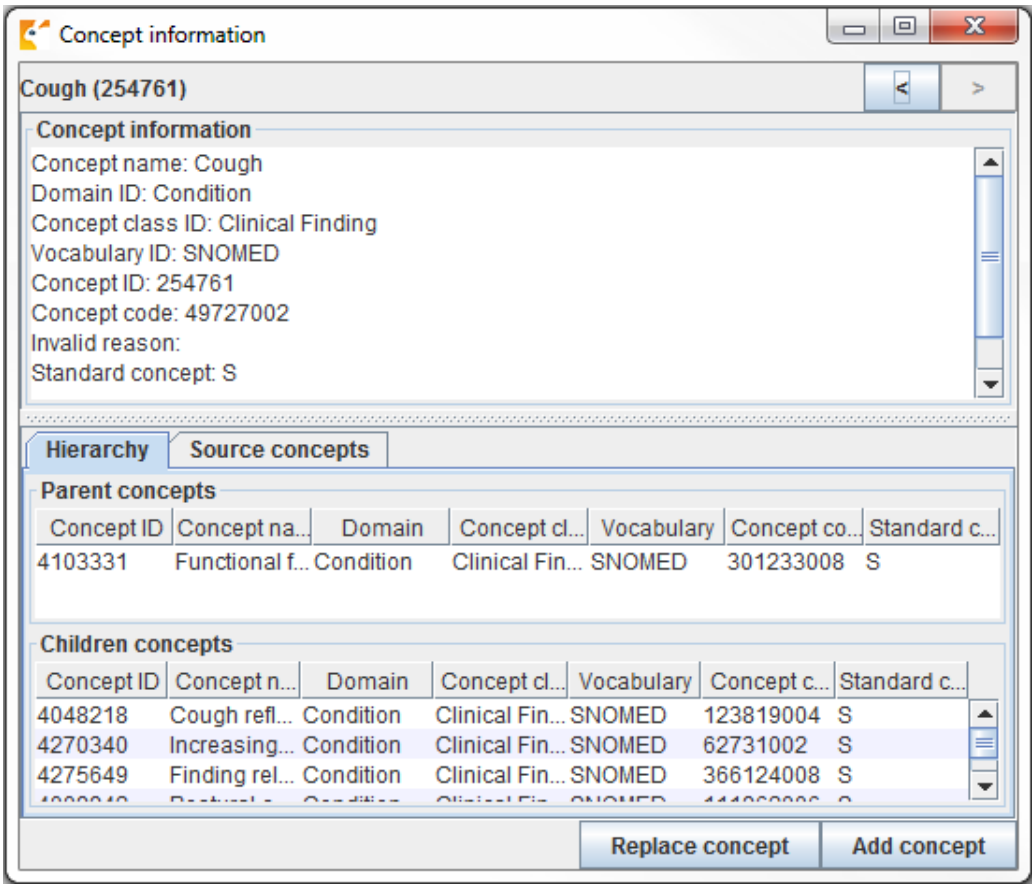


Figure 6.9: Usagi □□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□File (□□□□) -> Export source_to_concept_map (□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□) □□□□□□□□□□SOURCE_VOCABULARY_I (□□□) □□□□□□□□□□Usagi□□□ID□SOURCE_VOCABULARY_ID□□□□□□□□SOURCE_TO

SOURCE_VOCABULARY_ID□□□□□□□□□□CSV□□□□□□□□□□□□□□□□□□□□CSV□□□□ (□□)□□□□□□□□□□□□□□□□□□CSV□□□□□□□□□□□□□□□□□□□

- ETL□□□□□□□□□

- 80/20□□□

□	□

□□ **6.3.** VISIT_OCCURRENCE□□□□□□□□□□□□□□□□□□□□Synthea□□□□□□□□□□□□□□□□□□□

Part III

Chapter 7

□□□□: *David Madigan*

OHDSI □□□□□

- □□□□□□□□□...□
- □□□□□□□□□□□□□□□...□
- □□□□□□□□□
- □□□□□□□□□□□□□□□□□□□□□□□□
- □□□□□□□
- □□□□□□□□□□□□
- X□□□□□□□□□
- □□□□□□□□
- □□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□

- □□□□□□
- □□□□□□
- □□□□□
- □□□□□□□
- □□□□□
- □□
- □□
- □□□□□□
- □□□□□□

7.3

□□□□□□□□□□□□□□□□□□□

11□□□□□□□□□

7.7

□□ 7.1. □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

1. □□□□□□□□□□□□□□□□NSAID□□□□□□□□□□□□□□□□□□□□GI□□□□□□□□□□□□□□
2. □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□GI□□□□□□□□□□□□□□□□□□
3. □□□□□□□□□□□□□□□□□□□□□□□□□□□□GI□□□□□□□□□□□□□□□□□□

□□ 7.2. □□□□□□□□□□□□GI□□□□□□□□□□□□□□□□□□□

Chapter 8

– – OHDSI

□□□□: *Martijn Schuemie & Frank DeFalco*

OHDSI□□□□

Library□□□□□

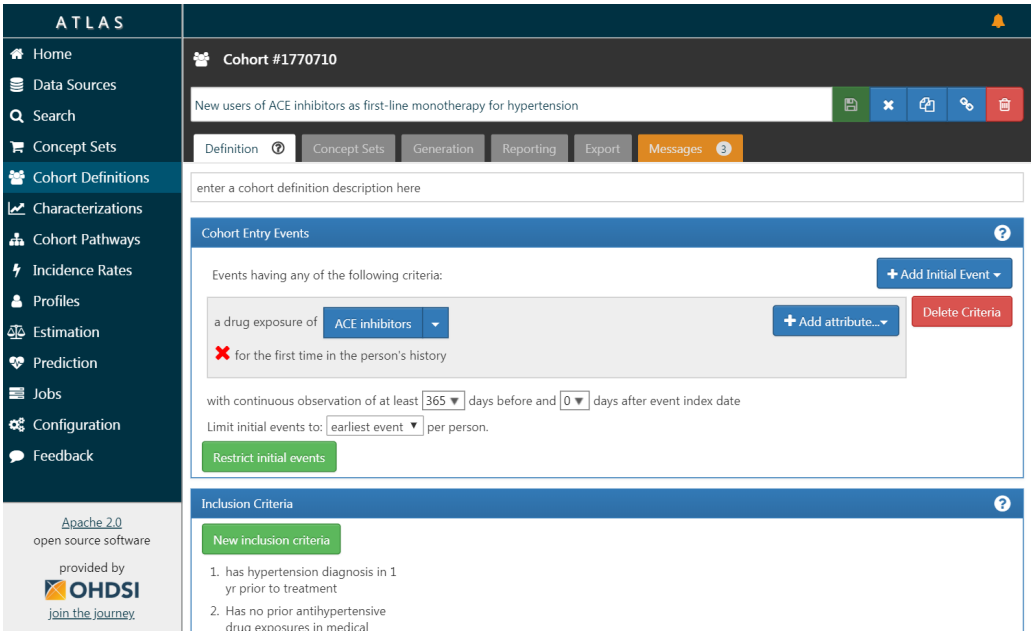


Figure 8.3: ATLAS□□□□□□□□□□□

□□□□□□□□ ATLAS□OMOP□□□□□□□□□□□

[2]□

8.3.2

ATLAS□□□□□□□□□□ATLAS GitHub□□□□□□wiki□□□□□□□□□□□□□□□ [3]□
□□wiki□□□□□□□□□□□□□□□

Figure 8.4: OHDSI Methods Library

2. □□□□□□

R□□□□□□□

1. https://cran.r-project.org/ □□□ 8.5 □□□□□□□□□□Download R for Windows □□□base□ □□□□□□□□□□□□□□□□□□□□□□□□

Figure 8.5: CRAN□□□R□□□□□□□

2. □□□□□□□□□□□□□□□□□□□□□□□□□□□□2□□□□□□□□□□□□□□□□□□□

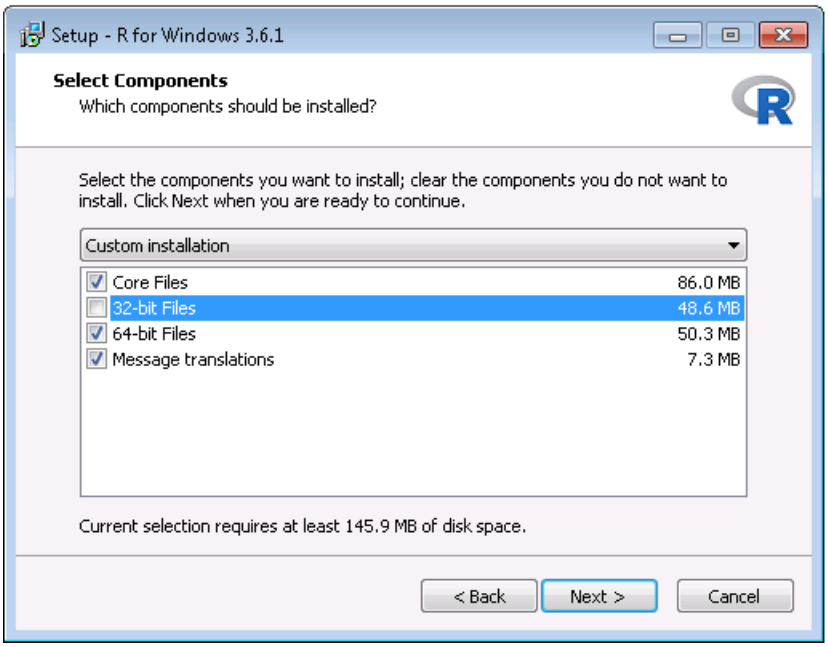


Figure 8.7: 32□□□□□□□□□□R□□□□

2. □□□□□□□□□□□

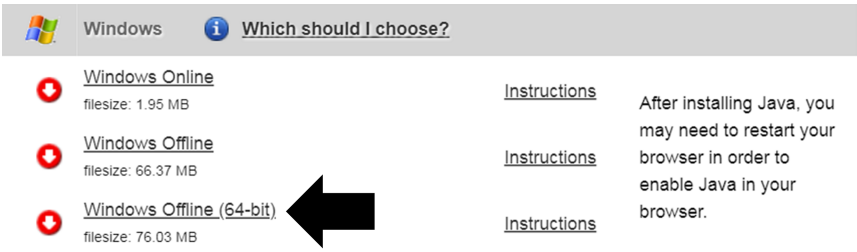


Figure 8.9: Java□□□□□□□

□□□□□□□□□

□□□□□□□□□

8.5.2 Amazon AWS

Amazon□□AWS□□□□□

8.6

- CDM□□□□□□□□□□□□□□□□□□
  - □□□□□□□□□□□□□
  - OHDSI Methods Library□R□□□□□□□□□□□□□□□□□□□□
  - □□□□□□□□□□□□□□□□□□□□□□ATLAS□□□□□□
- OHDSI□□□□□□□□□□□□□□□□□□□□□□□
  - □□□□
  - □□□□□□□□□□
  - □□□□□□□□□□□
- OHDSI□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
  - □□□□□□□□□□□□□□□□□□□□□□ATLAS
  - OHDSI Methods Library□R□□□□□□
- OHDSI□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

Chapter 9

– – SQL R

□□: *Martijn Schuemie & Peter Rijnbeek*

□□□□□□□□□□CDM□□□□□□□□□□

9.1 SqlRender

SqlRender □□□□□□ CRAN□Comprehensive R Archive Network□□□□□

```
sql <- "SELECT * FROM cohort {@x} ? {WHERE subject_id = 1}"
render(sql, x = FALSE)
```

```
## [1] "SELECT * FROM cohort "
```

```
render(sql, x = TRUE)
```

```
## [1] "SELECT * FROM cohort WHERE subject_id = 1"
```

□□□□□□□□□□□□□□□□□□

```
sql <- "SELECT * FROM cohort {@x == 1} ? {WHERE subject_id = 1};"
render(sql, x = 1)
```

```
## [1] "SELECT * FROM cohort WHERE subject_id = 1;"
```

```
render(sql, x = 2)
```

```
## [1] "SELECT * FROM cohort ;"
```

IN □□□□□□□□□□□□□□□□□

```
sql <- "SELECT * FROM cohort {@x IN (1,2,3)} ? {WHERE subject_id = 1};"
render(sql, x = 2)
```

```
## [1] "SELECT * FROM cohort WHERE subject_id = 1;"
```

9.1.2 SQL

SqlRender □□□□□□□□□□□□□□□□□OHDSI SQL□□□□□SQL□□□□□□□□□□□□□□□□□□

```
sql <- "SELECT TOP 10 * FROM person;"
translate(sql, targetDialect = "postgresql")
```

```
## [1] "SELECT  * FROM person LIMIT 10;"
## attr(,"sqlDialect")
## [1] "postgresql"
```

`targetDialect` □□□□□□□□□□□□□□□□□□□□□“oracle”, “postgresql”, “pdw”, “redshift”, “impala”, “netezza”, “bigquery”, “sqlite”, “sql server”□

SQL□□□□□

```sql
-- Nested queries:
SELECT * FROM (SELECT * FROM table_1) tmp WHERE a = b;

-- Limiting to top rows:
SELECT TOP 10 * FROM table;

-- Selecting into a new table:
SELECT * INTO new_table FROM table;

-- Creating tables:
CREATE TABLE table (field INT);

-- Inserting verbatim values:
INSERT INTO other_table (field_1) VALUES (1);

-- Inserting from SELECT:
INSERT INTO other_table (field_1) SELECT value FROM table;

-- Simple drop commands:
DROP TABLE table;

-- Drop table if it exists:
IF OBJECT_ID('ACHILLES_analysis', 'U') IS NOT NULL
  DROP TABLE ACHILLES_analysis;

-- Drop temp table if it exists:
IF OBJECT_ID('tempdb..#cohorts', 'U') IS NOT NULL
  DROP TABLE #cohorts;

-- Common table expressions:
WITH cte AS (SELECT * FROM table) SELECT * FROM cte;

-- OVER clauses:
SELECT ROW_NUMBER() OVER (PARTITION BY a ORDER BY b)
  AS "Row Number" FROM table;

-- CASE WHEN clauses:
SELECT CASE WHEN a=1 THEN a ELSE 0 END AS value FROM table;

-- UNIONs:
SELECT * FROM a UNION SELECT * FROM b;

-- INTERSECTIONs:
SELECT * FROM a INTERSECT SELECT * FROM b;

-- EXCEPT:
SELECT * FROM a EXCEPT SELECT * FROM b;
```

□□□□□□

□□□□□□□□□SQL Server□□□□□□□□□□□□□□□□□□□□□1□□□□□SQL Server□□□SELECT first_name + ' ' + last_name AS full_name FROM table □□□□□□□□□□□ PostgreSQL □ Oracle □□ SELECT first_name || ' ' || last_name AS full_name FROM table □□□□□□□□□□□□□SqlRender□□□□□□□□□□□□□□□□□□□

```
## Warning: The 'oracleTempSchema' argument is deprecated. Use 'tempEmulatio
## This warning is displayed once every 8 hours.
```

```
## [1] "SELECT * FROM temp_schema.yf8ly2lcchildren ;"
## attr(,"sqlDialect")
## [1] "oracle"
```

□□□□□ temp_schema □□□□□□□□□□□□□□□□□□□□□□□□

□□□Oracle□□□□□□□□□□□□□30□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□

SQL Server□□cdm_data.dbo.
□ cdm_data □□□□□□□□□ dbo □□□□□□□ person □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
Server□□□□□□□□□□□□□□□□□1□□□□□□□□□dbo □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
Server□□□□□□□□□□□□□□□□□□□

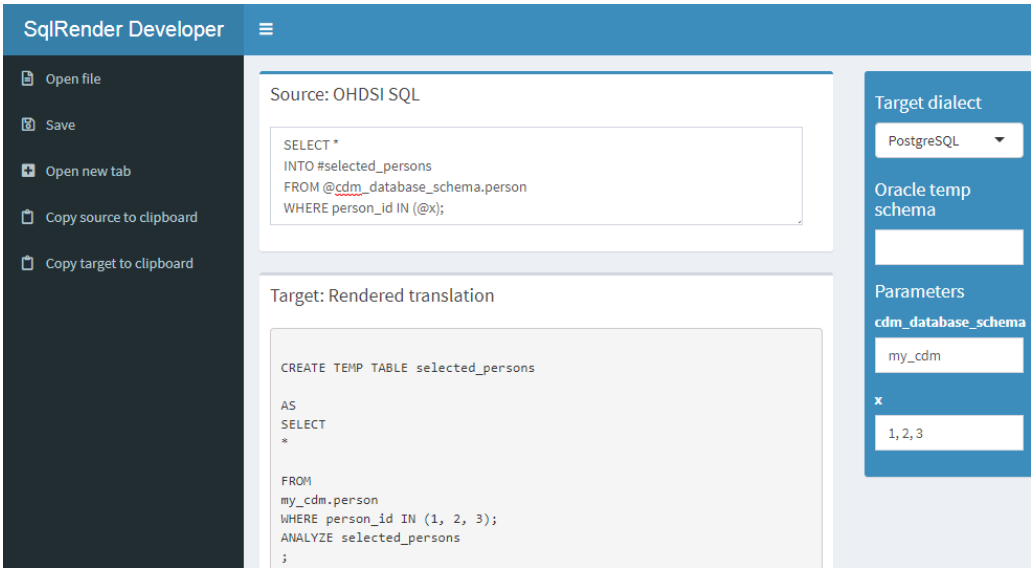


Figure 9.1: The SqlDeveloper Shiny app.

9.2.1

□□□

□□□□□

```
x <- renderTranslateQuerySql(conn,
                             sql = "SELECT TOP 10 * FROM @schema.person",
                             schema = "cdm_synpuf")
```

SQL Server□□□□TOP 10□□□□□□PostgreSQL□□□LIMIT 10□□□□□□□□□□SQL□□□□□□□

9.2.5

□□□□□□□□□□□□□□□□□□□executeSql□□□□□□□□SQL□□□□□□□□□□□□□□□□□□□□□□

```
data(mtcars)
insertTable(conn, "mtcars", mtcars, createTable = TRUE)
```

□□□□□□mtcars□□□□□□□□□□□□□□□□□□mtcars□□□□□□□□□□□□□□□□□□□□□□□□□□□□

9.3 CDM

□□□□□□□□OHDSI SQL□□□□□□CDM□□□□□□□□□□□□□□□□□□□

```sql
SELECT MAX(YEAR(observation_period_end_date) -
           year_of_birth) AS max_age
FROM @cdm.person
INNER JOIN @cdm.observation_period
  ON person.person_id = observation_period.person_id;
```

MAX_AGE
90

□□□□□

```
    SELECT COUNT(*) AS n
    FROM ages
    ) population_size;
```

MIN_AGE	Q25_AGE	MEDIAN_AGE	Q75_AGE	MAX_AGE
0	6	17	34	90

□□□□□□□□□□SQL□□□□□□R□□R□□

```
sql <- "SELECT YEAR(observation_period_start_date) -
               year_of_birth AS age
FROM @cdm.person
INNER JOIN @cdm.observation_period
  ON person.person_id = observation_period.person_id;"
age <- renderTranslateQuerySql(conn, sql, cdm = "cdm")
quantile(age[, 1], c(0, 0.25, 0.5, 0.75, 1))
```

```
##   0%  25%  50%  75% 100%
##    0    6   17   34   90
```

□□□□□□□□□□□□□□□□□□□

9.4

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□CDM□□□□□□□□SQL□□□□□□□□□□□□

```
SELECT COUNT(*) AS subject_count,
  concept_name
FROM @cdm.person
INNER JOIN @cdm.concept
  ON person.gender_concept_id = concept.concept_id
GROUP BY concept_name;
```

SUBJECT_COUNT	CONCEPT_NAME
14927548	FEMALE
11371453	MALE

□□□□□□□□□□□□□□□□□□□

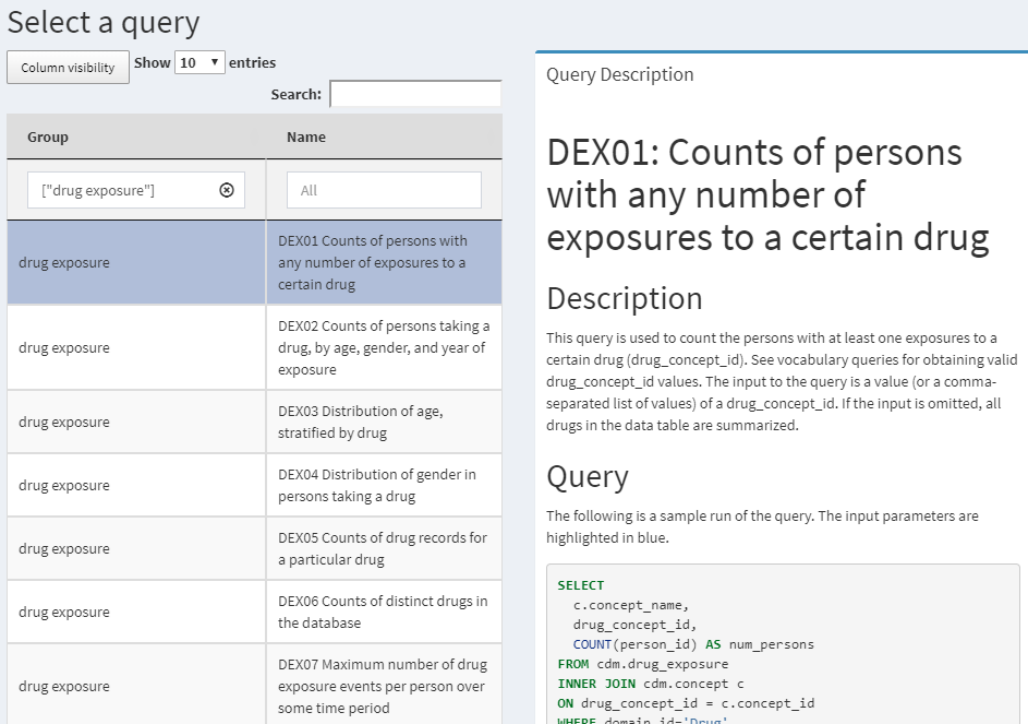


Figure 9.2: □□□□□□□□□□CDM□□□□SQL□□□□□□□□□□□

9.6

9.6.1

□□□□□□□ACE□□□□ACEi□□□□□□□□□□□□□□□Slater et al. (1988) □□□□□ACEi□□□□□1□□□□□□□□□□□□□□3,000□□1□/□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□1□□□□□□□□□□□□□□□□□□□

9.7 SQL R

OHDSI□□□□□

```
      drug_era_end_date,
      person_id,
      ROW_NUMBER() OVER (
        PARTITION BY person_id
            ORDER BY drug_era_start_date
      ) order_nr
    FROM @cdm_db_schema.drug_era
    WHERE drug_concept_id = 1308216 --
  ) ordered_exposures
  WHERE order_nr = 1
) first_era
INNER JOIN @cdm_db_schema.observation_period
  ON subject_id = person_id
    AND observation_period_start_date < cohort_start_date
    AND observation_period_end_date > cohort_start_date
WHERE DATEDIFF(DAY,
               observation_period_start_date,
               cohort_start_date) >= 365;
"

renderTranslateExecuteSql(conn, sql,
                          cohort_db_schema = cohortDbSchema,
                          cohort_table = cohortTable,
                          cdm_db_schema = cdmDbSchema)
```

□□□□□CDM□□□□□□□□□□□DRUG_ERA□□□□□□□□□□□□□□□□□□□□DRUG_EXPOS
□□□□□□□□□□1□□□□□□□□□□□□□□□□□□□□□□1□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□ COHORT_START_DATE □□□□□□□□□□□ 365 □□□□□□□□□□□□□□□

9.7.2

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

```
sql <- "
INSERT INTO @cohort_db_schema.@cohort_table (
 cohort_definition_id,
 cohort_start_date,
 cohort_end_date,
subject_id
)
SELECT 2 AS cohort_definition_id,
  cohort_start_date,
  cohort_end_date,
  subject_id
FROM (
  SELECT DISTINCT person_id AS subject_id,
    condition_start_date AS cohort_start_date,
```

```
    condition_end_date AS cohort_end_date
  FROM @cdm_db_schema.condition_occurrence
  INNER JOIN @cdm_db_schema.concept_ancestor
    ON condition_concept_id = descendant_concept_id
  WHERE ancestor_concept_id = 432791 --
) distinct_occurrence
INNER JOIN @cdm_db_schema.visit_occurrence
  ON subject_id = person_id
  AND visit_start_date <= cohort_start_date
  AND visit_end_date >= cohort_start_date
WHERE visit_concept_id IN (262, 9203,
    9201) --    ER;
"

renderTranslateExecuteSql(conn, sql,
                          cohort_db_schema = cohortDbSchema,
                          cohort_table = cohortTable,
                          cdm_db_schema = cdmDbSchema)
```

□□□□□CONDITION_OCCURRENCE□□□□□CONCEPT_ANCESTOR□□□□□

```
SELECT days.gender,
    days.age,
    days,
    CASE WHEN events IS NULL THEN 0 ELSE events END AS events
FROM (
  SELECT gender,
    age,
    SUM(DATEDIFF(DAY, cohort_start_date,
      cohort_end_date)) AS days
  FROM tar
  GROUP BY gender,
    age
) days
LEFT JOIN (
  SELECT gender,
      age,
      COUNT(*) AS events
  FROM tar
  INNER JOIN @cohort_db_schema.@cohort_table angioedema
    ON tar.subject_id = angioedema.subject_id
      AND tar.cohort_start_date <= angioedema.cohort_start_date
      AND tar.cohort_end_date >= angioedema.cohort_start_date
  WHERE cohort_definition_id = 2 --
  GROUP BY gender,
    age
) events
ON days.gender = events.gender
  AND days.age = events.age;
"

results <- renderTranslateQuerySql(conn, sql,
                                   cohort_db_schema = cohortDbSchema,
                                   cohort_table = cohortTable,
                                   cdm_db_schema = cdmDbSchema,
                                   snakeCaseToCamelCase = TRUE)
```

□□□CTE□tar□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□OBSERVATION_PERIOD_

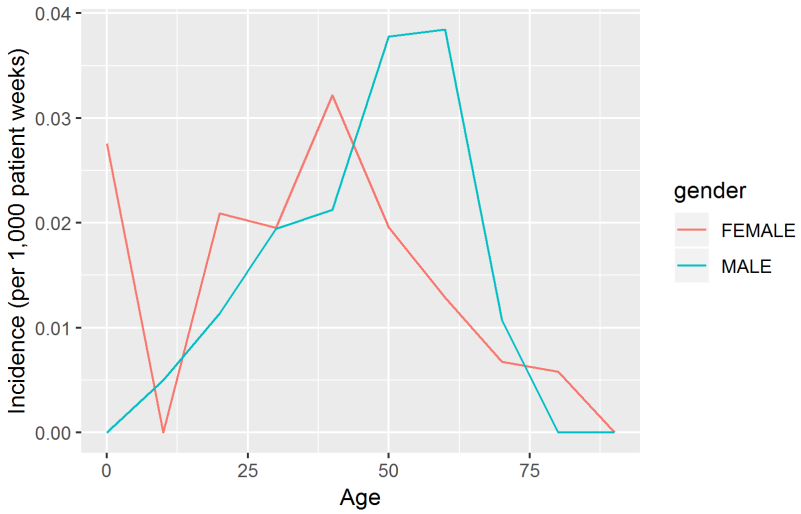
snakeCaseToCamelCase = TRUE □□□□□□□SQL□□□□□□□□□□snake_case □□□□□□□□□□□□SQL□□□□□□□□□□□□□□□□□□□□□□□□□R□□camelCase□□□□□□□

ggplot2□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

```
#   IR
results$ir <- 1000 * results$events / results$days / 7

#
results$age <- results$age * 10
```

```
library(ggplot2)
ggplot(results, aes(x = age, y = ir, group = gender, color = gender)) +
  geom_line() +
  xlab(" ") +
  ylab("  1,000 / ")
```



9.7.4

□□□□□

9.8

- **SQL**□Structured Query Language□□□□□□□□□□□□□CDM□□□□□□□□□□□□□
- □□□□□□□□□□□□□□□□□□□□□□□SQL□□□□□□□□□□□□□□□□□□□□□□□□□
- **SqlRender**□**DatabaseConnector**R□□□□□□□□CDM□□□□□□□□□□□□□□□□□□□
- R□SQL□□□□□□□□□□□OHDSI□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
- **QueryLibrary**□□CDM□□□□□□□□□SQL□□□□□□□□□□□□□□□□□□□□

9.9

□□□□

□□□□□□□□□□□□□□□□ 8.4.5 □□□□□□□□□□□□□□□R□R-Studio□Java □□□□□□□□□□□□□□□
Eunomia □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

```
install.packages(c("SqlRender", "DatabaseConnector", "remotes"))
remotes::install_github("ohdsi/Eunomia", ref = "v1.0.0")
```

Eunomia □□□□□□□□CDM □□□□□□□ R □□□

```
connectionDetails <- Eunomia::getEunomiaConnectionDetails()
```

CDM □□□□□□□□□□□□□□□main□□□□

□□ **9.1.** SQL □ R □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□ **9.2.** SQL □ R □□□□□□□□□□□□□□□□□□□□

Chapter 10

— —

□□□□: *Kristin Kostka*

□□□□

10.2

□□□□□□□

10.3

□□□□□□□□

10.5

□□□□□□□

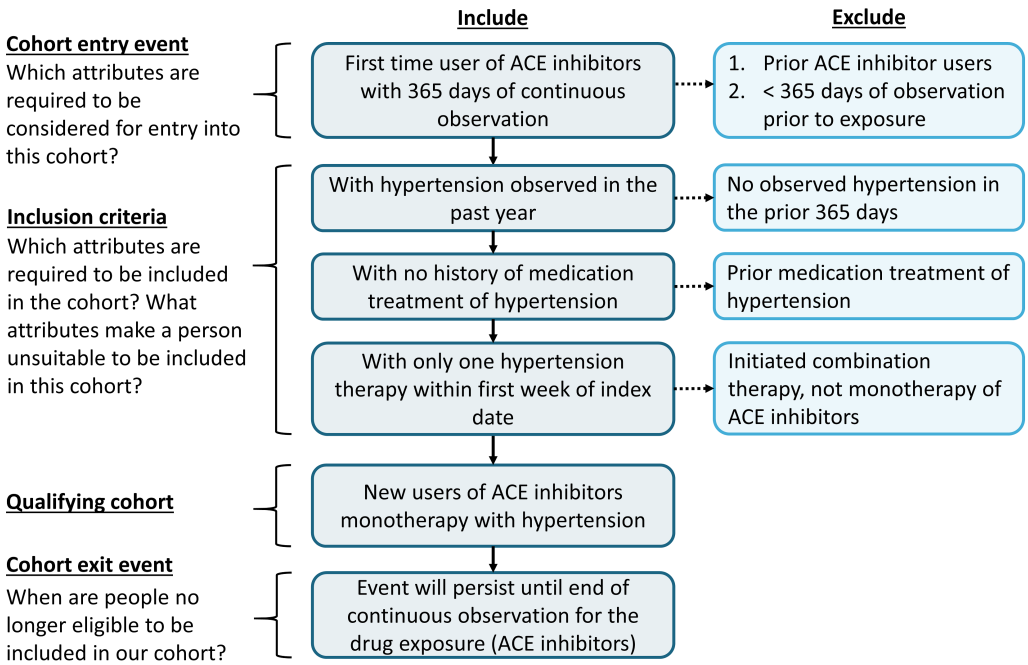


Figure 10.3: □□□□□□□□□□□□□□

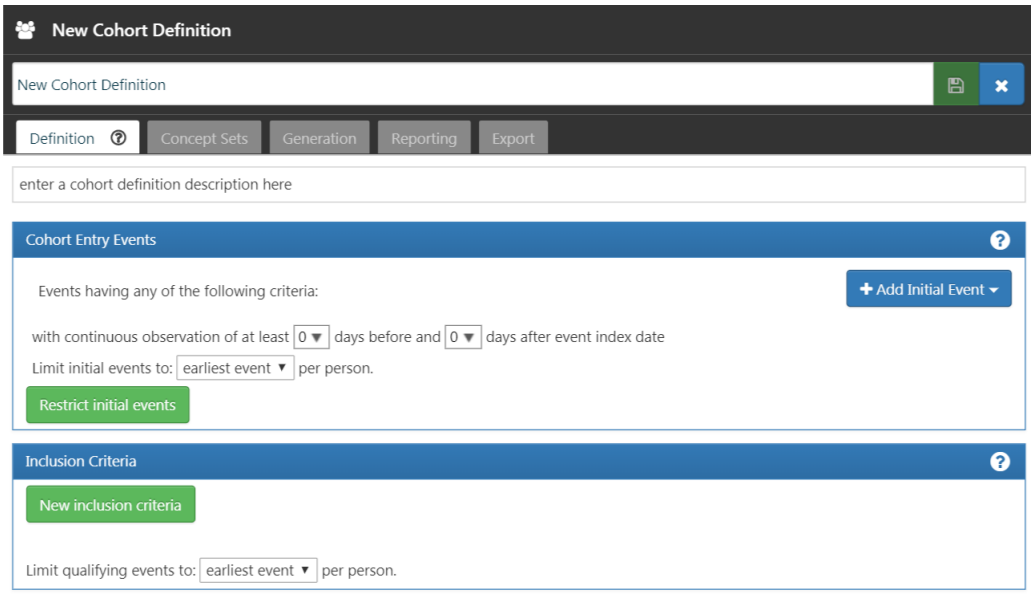


Figure 10.4: □□□□□□□□□□

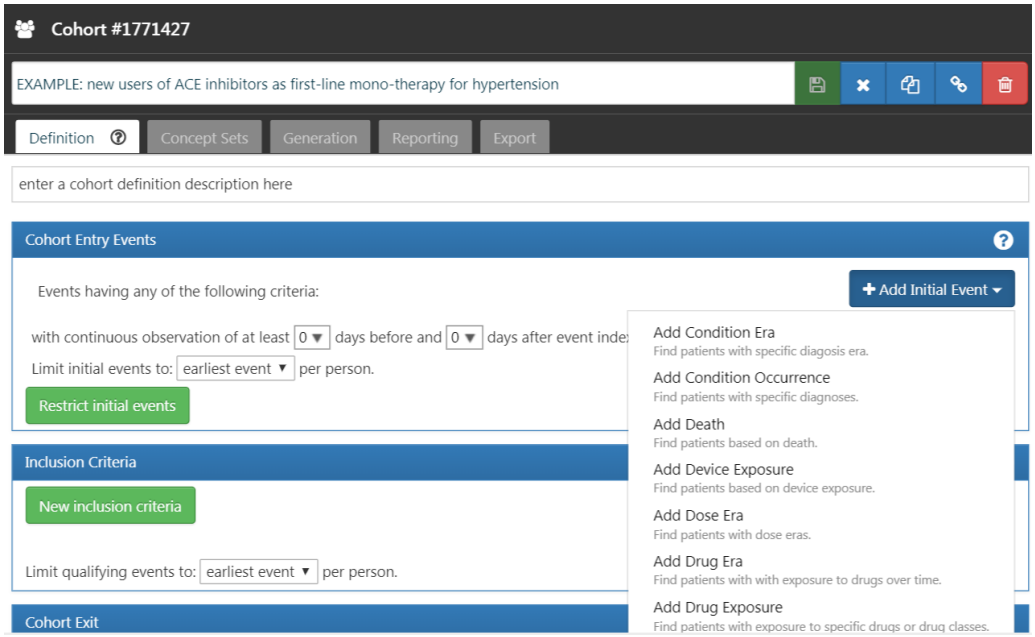


Figure 10.5: □□□□□□□□□□

□□□10.6□□□□□□□□□ATLAS□□□

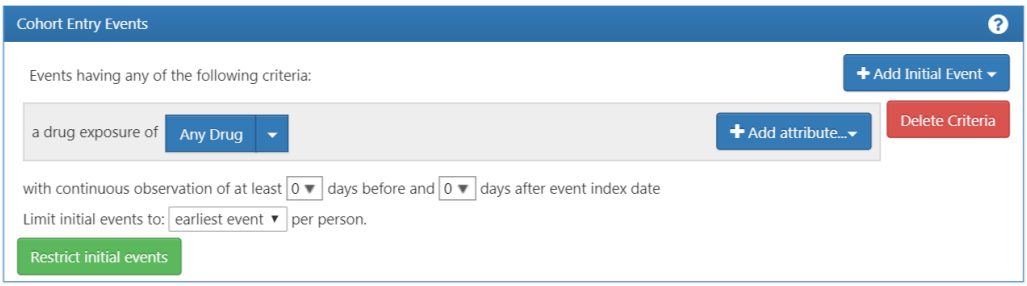


Figure 10.6: □□□□□□□□

10.7.2

□□□□□□□□□□□□□□□□□□□

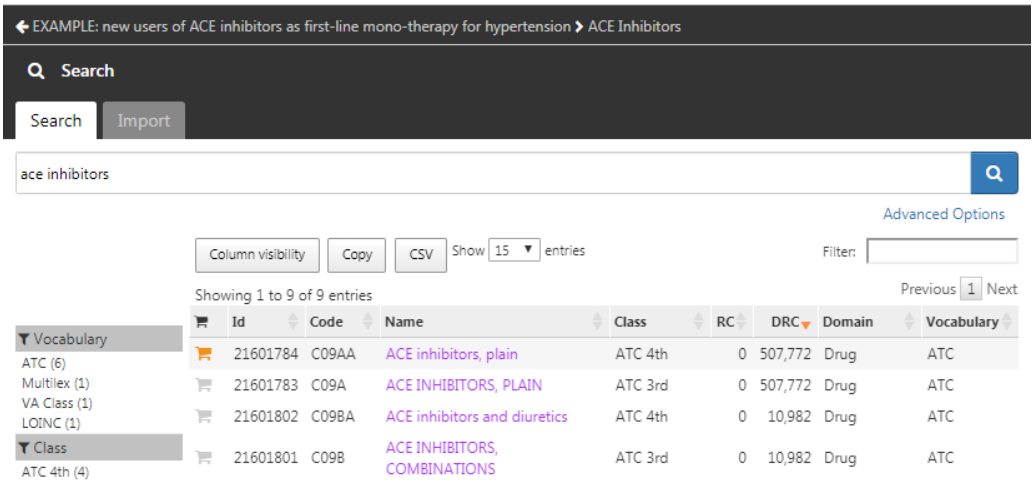


Figure 10.7: □□□□□ - ACE□□□

□10.8□□□□□□□□□□□□

Concept Set Expression | Included Concepts 21536 | Included Source Codes | Export | Import

Name:

ACE Inhibitors

Show 25 entries Search:

Showing 1 to 15 of 15 entries Previous 1 Next

Concept Id	Concept Code	Concept Name	Domain	Standard Concept Caption	Exclude	Descendants	Mapped
1335471	18867	benazepril	Drug	Standard		✓	
1340128	1998	Captopril	Drug	Standard		✓	
19050216	21102	Cilazapril	Drug	Standard		✓	
1341927	3827	Enalapril	Drug	Standard		✓	
1342001	3829	Enalaprilat	Drug	Standard		✓	
1363749	50166	Fosinopril	Drug	Standard		✓	
19122327	60245	imidapril	Drug	Standard		✓	
1308216	29046	Lisinopril	Drug	Standard		✓	
1310756	30131	moexipril	Drug	Standard		✓	
1373225	54552	Perindopril	Drug	Standard		✓	
1331235	35208	quinapril	Drug	Standard		✓	
1334456	35296	Ramipril	Drug	Standard		✓	
19040051	36908	spirapril	Drug	Standard		✓	
1342439	38454	trandolapril	Drug	Standard		✓	
19102107	39990	zofenopril	Drug	Standard		✓	

Classification Non-Standard Standard

Figure 10.8: ACE□□□□□□□□□□□□□□□□

Import Concept Set From Repository...

New Concept Set

Show 10 entries Filter Repository Concept Sets: ace inhibitors

Id	Title	Created	Modified	Author
1794480	[OHDSI EU 2019] Excluded concepts of ACE inhibitors or Thiazide diuretics	03/28/2019 11:04 AM	03/28/2019 11:04 AM	anonymous
963	ACE Inhibitors			anonymous
3268	COPY OF: ACE Inhibitors			anonymous
99283	Ace Inhibitors			anonymous
142965	PheKB ACE-I ACE inhibitors			anonymous

Showing 1 to 5 of 5 entries (filtered from 11,667 total entries) Previous 1 Next

Figure 10.9: ATLAS□□□□□□□□□□□□□□□□

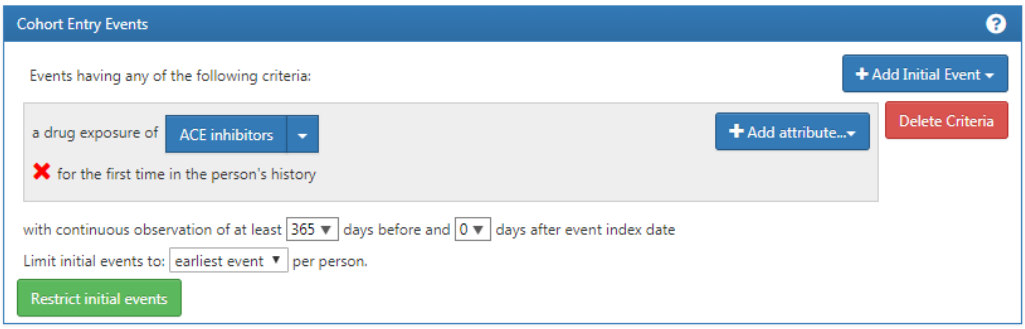


Figure 10.10: □□□□□□□□□□□□□□□□□□□□□□□□

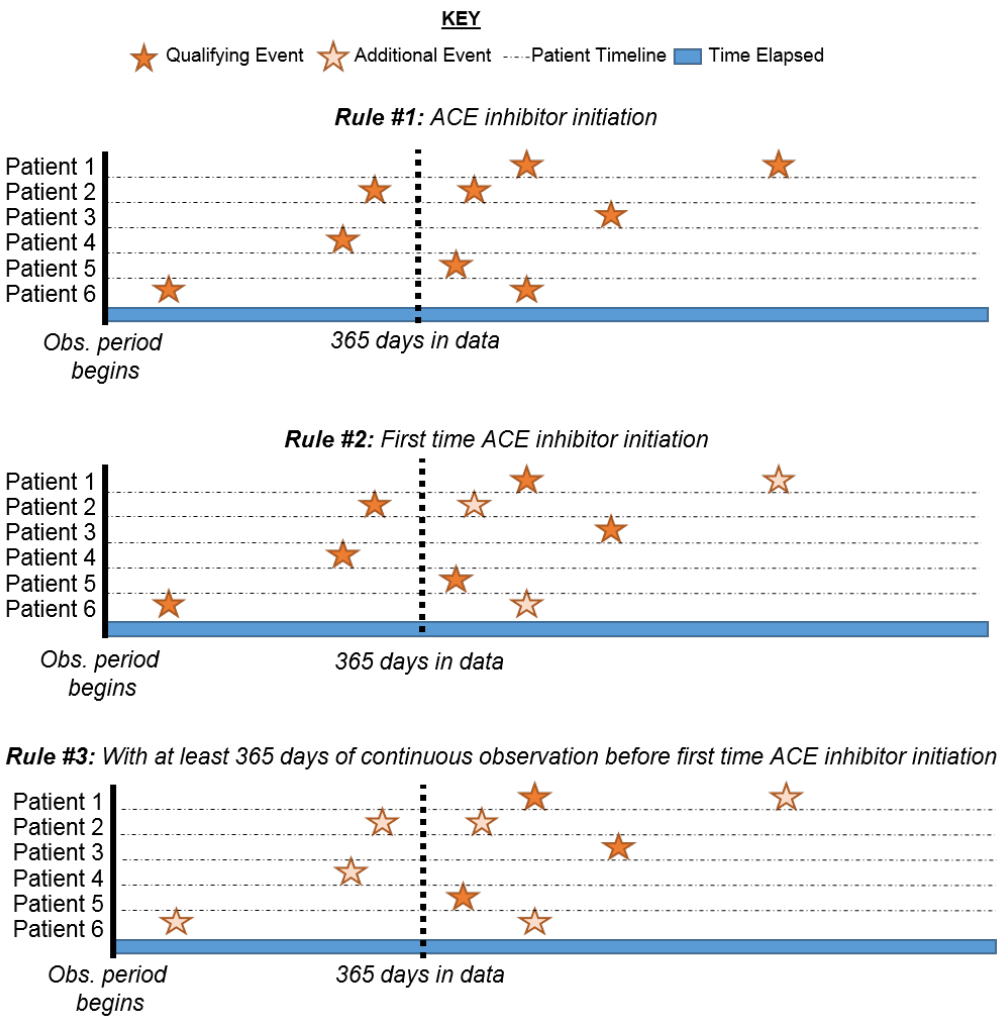


Figure 10.11: □□□□□□□□□□□□□□□□□□□□□

10.7.4

□□□□□□□

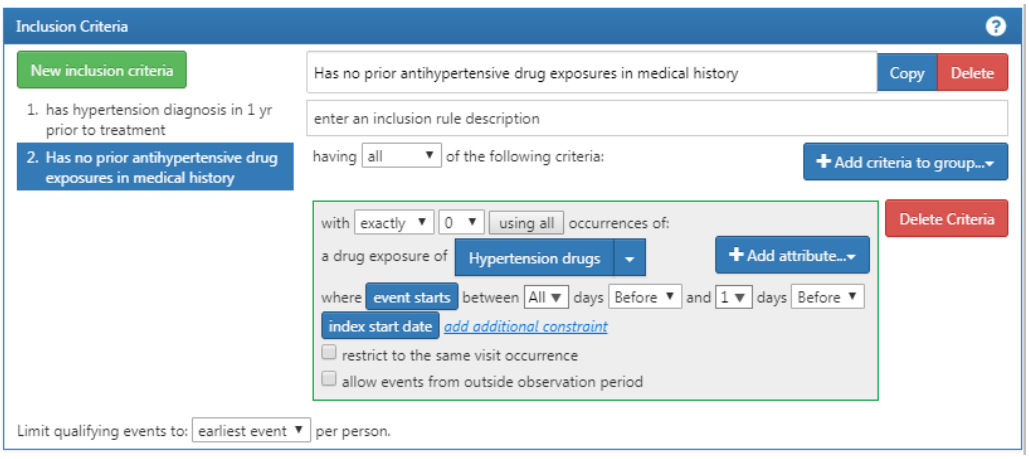


Figure 10.13: □□□□□□□ 2

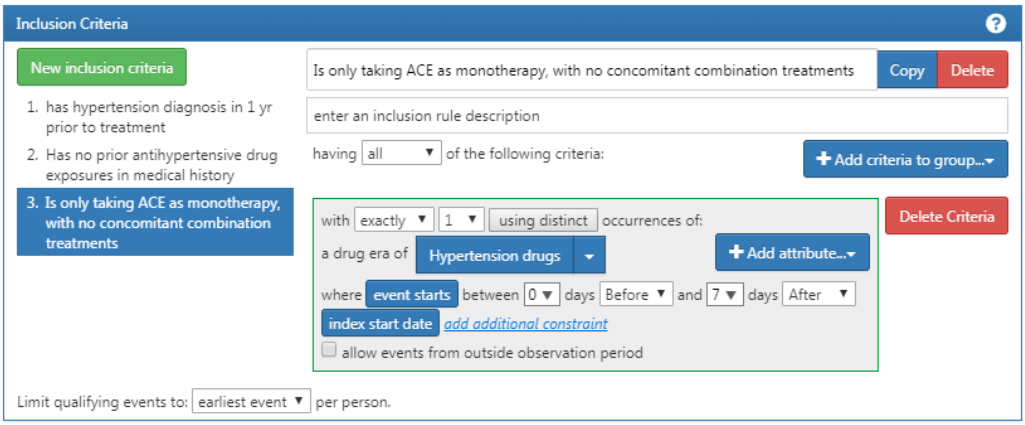


Figure 10.14: □□□□□□□ 3

10.7.5

□□□□□□□

```
cdmDbSchema <- "my_cdm_data"
cohortDbSchema <- "scratch"
cohortTable <- "my_cohorts"
```

□□□3□□□cdmDbSchema□cohortDbSchema□□□□cohortTable□□□□□□□□□□□□□

```
                          cdm_db_schema = cdmDbSchema,
                          ace_i = aceI)
```

DRUG_EXPOSURE□□□□□□CONCEPT_ANCESTOR□□□□□□□□□□□□□□□□ACE□□□□□□□□□□□

10.8.4 365

□□□OBSERVATION_PERIOD□□□□□□□□□□365□□□□□□□□□□□□□□□□□□□□□

```
sql <- "SELECT subject_id,
  cohort_start_date
INTO #has_prior_obs
FROM #first_use
INNER JOIN @cdm_db_schema.observation_period
  ON subject_id = person_id
    AND observation_period_start_date <= cohort_start_date
    AND observation_period_end_date >= cohort_start_date
WHERE DATEADD(DAY, 365, observation_period_start_date) < cohort_start_date;"

renderTranslateExecuteSql(conn, sql, cdm_db_schema = cdmDbSchema)
```

10.8.5

365□□□□□□□□□□□□□□□□□

```
sql <- "SELECT DISTINCT subject_id,
  cohort_start_date
INTO #has_ht
FROM #has_prior_obs
INNER JOIN @cdm_db_schema.condition_occurrence
  ON subject_id = person_id
    AND condition_start_date <= cohort_start_date
    AND condition_start_date >= DATEADD(DAY, -365, cohort_start_date)
INNER JOIN @cdm_db_schema.concept_ancestor
  ON descendant_concept_id = condition_concept_id
WHERE ancestor_concept_id = @hypertension;"

renderTranslateExecuteSql(conn,
                          sql,
                          cdm_db_schema = cdmDbSchema,
                          hypertension = hypertension)
```

□□SELECT DISTINCT□□□□□□□□□□□□□□□□□□□□□

10.8.6

□□□□□□□□□□□□□□□□□□□□□□□□

```
sql <- "SELECT subject_id,
  cohort_start_date
INTO #no_prior_ht_drugs
FROM #has_ht
LEFT JOIN (
  SELECT *
  FROM @cdm_db_schema.drug_exposure
  INNER JOIN @cdm_db_schema.concept_ancestor
    ON descendant_concept_id = drug_concept_id
  WHERE ancestor_concept_id IN (@all_ht_drugs)
) ht_drugs
  ON subject_id = person_id
    AND drug_exposure_start_date < cohort_start_date
WHERE person_id IS NULL;"

renderTranslateExecuteSql(conn,
                          sql,
                          cdm_db_schema = cdmDbSchema,
                          all_ht_drugs = allHtDrugs)
```

LEFT JOIN□□□□□□□□□DRUG_EXPOSURE□□□□□□□□person_id□NULL□□□□□□□□□□□

10.8.7

□□□□□□□□□□□□□□7□□□□□□□□□□□□□□□□□□□□□□□□□

```
sql <- "SELECT subject_id,
  cohort_start_date
INTO #monotherapy
FROM #no_prior_ht_drugs
INNER JOIN @cdm_db_schema.drug_exposure
  ON subject_id = person_id
    AND drug_exposure_start_date >= cohort_start_date
    AND drug_exposure_start_date <= DATEADD(DAY, 7, cohort_start_date)
INNER JOIN @cdm_db_schema.concept_ancestor
  ON descendant_concept_id = drug_concept_id
WHERE ancestor_concept_id IN (@all_ht_drugs)
GROUP BY subject_id,
  cohort_start_date
HAVING COUNT(*) = 1;"

renderTranslateExecuteSql(conn,
                          sql,
                          cdm_db_schema = cdmDbSchema,
```

```
                                all_ht_drugs = allHtDrugs)
```

10.8.8

□□□□□□□

```
        MAX(start_ordinal) OVER (
          PARTITION BY person_id ,concept_id ORDER BY event_date,
              event_type ROWS UNBOUNDED PRECEDING
          ) AS start_ordinal,
        ROW_NUMBER() OVER (
          PARTITION BY person_id, concept_id ORDER BY event_date,
            event_type
          ) AS overall_ord
      FROM (
-- select the start dates, assigning a row number to each
        SELECT person_id,
          concept_id,
          exposure_start_date AS event_date,
          0 AS event_type,
          ROW_NUMBER() OVER (
            PARTITION BY person_id, concept_id ORDER BY exposure_start_date
            ) AS start_ordinal
        FROM #exposure exposure

        UNION ALL
-- add the end dates with NULL as the row number, padding the end dates by
-- @max_gap to allow a grace period for overlapping ranges.

        SELECT person_id,
          concept_id,
          DATEADD(day, @max_gap, exposure_end_date),
          1 AS event_type,
          NULL
        FROM #exposure exposure
        ) rawdata
    ) events
  WHERE 2 * events.start_ordinal - events.overall_ord = 0
  ) events
  ON exposure.person_id = events.person_id
      AND exposure.concept_id = events.concept_id
      AND events.end_date >= exposure.exposure_end_date
  GROUP BY exposure.person_id,
      exposure.concept_id,
      exposure.exposure_start_date
  ) ends
GROUP BY ends.person_id,
  concept_id,
  ends.era_end_date;"

renderTranslateExecuteSql(conn,
                          sql,
                          cdm_db_schema = cdmDbSchema,
                          max_gap = 30)
```

□□□□□□□□□□□□□□□□□□max_gap□□□□□□□□□□□□□□□□□□

10.9

- □□□□□□□□1□□□□□□□□□□□□□□□□□□□□□□□□□□□
- □□□
- □□□□□□□□□□□□□□□□□□□□

□□ **10.2.** SQL□□□R□□□□□□□□□□COHORT□□□□□□□□□□□□□□□AMI□□□□□□□□□□□□□□

- □□□□□□□□□□□□□□□□□□4329847□□□□□□□□□□□□□□□□□

Chapter 11

— —

□□□□: *Anthony Sena & Daniel Prieto-Alhambra*

□□□□

□□□□□

□ 11.2 □□□□□□□□□□□□

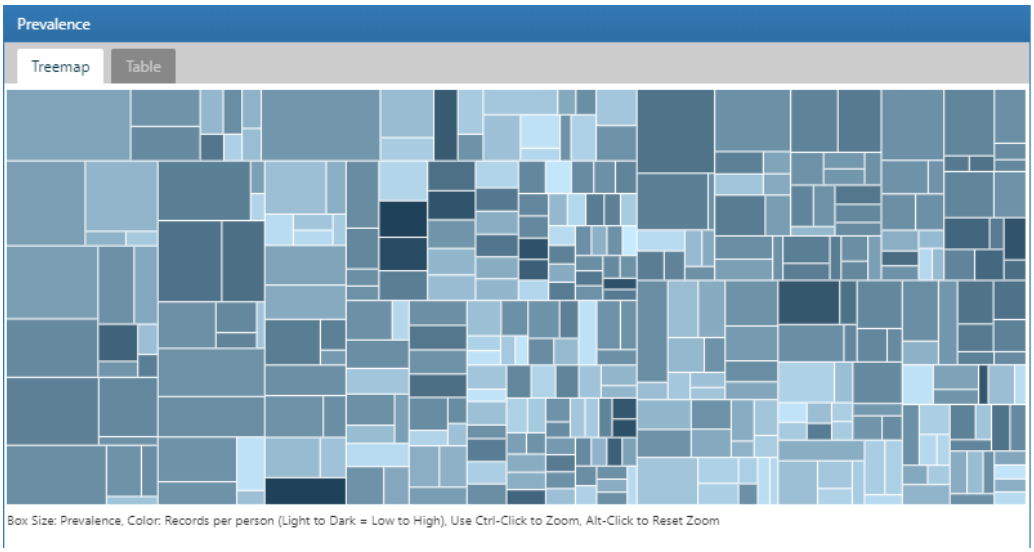


Figure 11.3: ATLAS□□□□□□: □□□□□□□□□□□□□□□□□

Prevalence

Treemap | Table

Column visibility | Copy | CSV | Show 15 entries | Filter: hypertension

Showing 1 to 15 of 47 entries
(filtered from 15,907 total entries)

Previous 1 2 3 4 Next

Concept Id	Name	Person Count	Prevalence	Records per person
320128	Essential hypertension	17,814,076	12.30%	5.80
312648	Benign essential hypertension	11,014,877	7.61%	4.35
317898	Malignant essential hypertension	1,021,441	0.70%	2.22
381290	Ocular hypertension	521,264	0.36%	2.40
441922	Transient hypertension of pregnancy	209,317	0.14%	2.45
44782429	Chronic kidney disease due to hypertension	170,534	0.12%	3.60
137940	Transient hypertension of pregnancy - delivered	153,806	0.11%	1.07
321080	Hypertension complicating pregnancy, childbirth and the puerperium	148,728	0.10%	2.15
314423	Benign essential hypertension complicating pregnancy, childbirth and the puerperium - not delivered	132,245	0.09%	3.94
44782690	Chronic kidney disease stage 5 due to hypertension	119,375	0.08%	5.20
44783618	Heritable pulmonary arterial hypertension	104,737	0.07%	3.61
319826	Secondary hypertension	96,356	0.07%	2.14
4167493	Pregnancy-induced hypertension	91,675	0.06%	2.60
321074	Pre-existing hypertension complicating pregnancy, childbirth and puerperium	74,311	0.05%	2.99
192680	Portal hypertension	71,240	0.05%	3.11

Showing 1 to 15 of 47 entries
(filtered from 15,907 total entries)

Previous 1 2 3 4 Next

Figure 11.4: ATLAS□□□□□□: □□□□□□□ ”hypertension (□□□)” □□□□□□□□□□□□

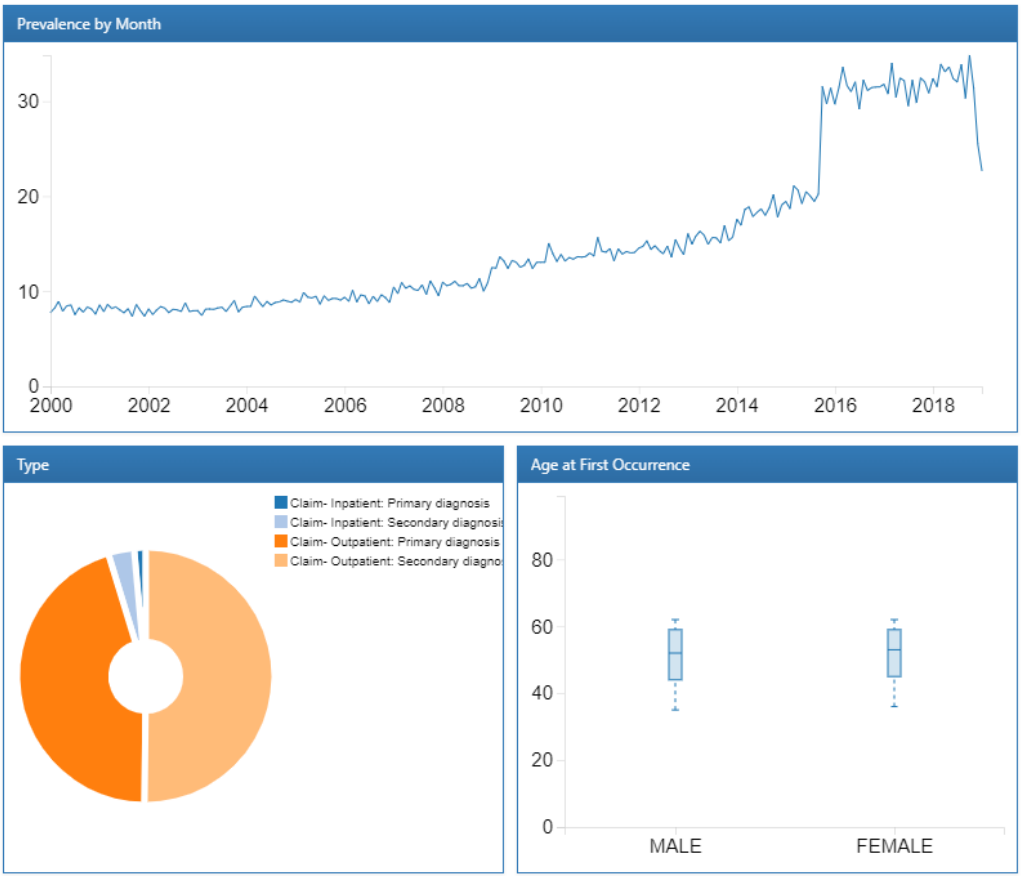


Figure 11.5: ATLAS□□□□□□: □□□□□□□□□□□□□□□□□□

□□□□□□

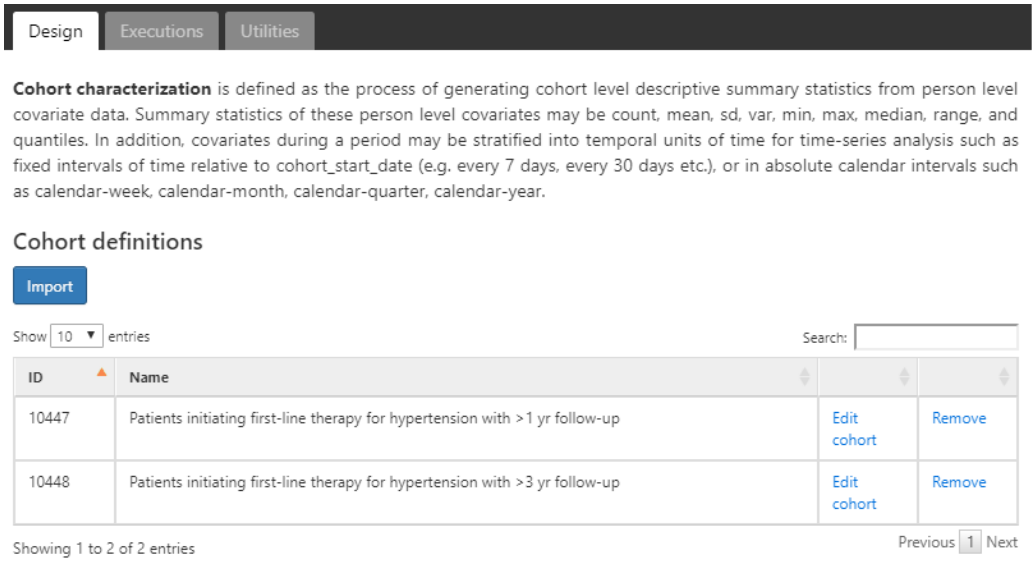


Figure 11.6: □□□□□□ - □□□□□□□□□□

□□□□□□□ATLAS□□□□□□□□□□□□□□□□□□□□□ 10 □□□□□□□ Import □□□□□□□□□
11.6 □□□□□□□□□□□□□□□□□□□□

Feature analyses

Import

Show 25 entries Search:

ID	Name	Description	Actions
43	Drug Era Short Term	One covariate per drug in the drug_era table overlapping with any part of the short window.	Remove
49	Charlson Index	The Charlson comorbidity index (Romano adaptation) using all conditions prior to the window end.	Remove
67	Condition Occurrence Long Term	One covariate per condition in the condition_occurrence table starting in the long term window.	Remove
71	Demographics Age Group	Age of the subject on the index date (in 5 year age groups)	Remove
72	Demographics Race	Race of the subject.	Remove
73	Demographics Prior Observation Time	Number of continuous days of observation time preceding the index date.	Remove
74	Demographics Gender	Gender of the subject.	Remove
76	Condition Occurrence Medium Term	One covariate per condition in the condition_occurrence table starting in the medium term window.	Remove
77	Demographics Age	Age of the subject on the index date (in years).	Remove
79	Demographics Time In Cohort	Number of days of observation time during cohort period.	Remove
80	Demographics Index Year	Year of the index date.	Remove
81	Demographics Post Observation Time	Number of continuous days of observation time following the index date.	Remove
87	Procedure Occurrence Any Time Prior	One covariate per procedure in the procedure_occurrence table any time prior to index.	Remove
103	Visit Count Long Term	The number of visits observed in the long term window.	Remove

Figure 11.7: □□□□□□ - □□□□

Subgroup analyses

New subgroup

Female

Calculate subgroup analyses only

Female

having all of the following criteria: Add criteria to group...

with the following event criteria:

Add attribute...

with a gender of: FEMALE Add Import

Delete Criteria

Figure 11.8: □□□□□□□□ - □□□□□□□□□□□

11.7.2

□□□□□□□□□□□□□□□□□□□□□□1□□□□□□□□□□□□□□□□□□□

Exploring condition_occurrence during day -365 through 0 days relative to index: Cardiac arrhythmia

Cohort: Patients initiating first-line therapy for hypertension with >1 yr follow-up

Export Show 10 entries Search:

Relationship type	Distance	Concept name	All stratas		Female	
			Count	Pct	Count	Pct
Explore Ancestor	4	Disorder by body site	32	0.00%	17	0.00%
Explore Ancestor	4	Finding of trunk structure	991	0.06%	605	0.08%
Explore Ancestor	3	Disorder of trunk	23	0.00%	14	0.00%
Explore Ancestor	3	Disorder of thorax	241	0.01%	104	0.01%
Explore Ancestor	3	Disorder of body system	4,135	0.25%	1,992	0.26%
Explore Ancestor	2	Disorder of cardiovascular system	12,979	0.78%	6,073	0.79%
Explore Ancestor	2	Disorder of mediastinum	138	0.01%	62	0.01%
Explore Ancestor	2	Disorder of body cavity	24	0.00%	10	0.00%
Explore Ancestor	1	Heart disease	4,691	0.28%	1,869	0.24%
Explore Selected	0	Cardiac arrhythmia	30,474	1.83%	13,221	1.72%

Showing 1 to 10 of 62 entries Previous 1 2 3 4 5 6 7 Next

Figure 11.11: □□□□□□□□ - □□□□□□□□□□□□

CONDITION / Condition Occurrence Long Term / stratified by Female

Export Export comparison Show 10 entries Search: edema

Covariate	Explore	Concept ID	Patients initiating first-line therapy for hypertension with >1 yr follow-up				Patients initiating first-line therapy for hypertension with >3 yr follow-up				Std diff
			Count	Pct	Female		Count	Pct	Female		
					Count	Pct			Count	Pct	
Edema	Explore	433595	32,243	1.94%	20,200	2.63%	15,173	1.81%	9,684	2.48%	-0.0066

Showing 1 to 1 of 1 entries (filtered from 206 total entries) Previous 1 Next

Figure 11.12: □□□□□□□□ - □□□□□□□□

Exploring condition_occurrence during day -365 through 0 days relative to index: Edema

Cohort: Patients initiating first-line therapy for hypertension with >1 yr follow-up

Export Show 10 entries Search: ang

Relationship type	Distance	Concept name	All stratas		Female	
			Count	Pct	Count	Pct
Explore Descendant	-2	Angioedema	2,605	0.16%	1,506	0.20%

Showing 1 to 1 of 1 entries (filtered from 56 total entries) Previous 1 Next

Figure 11.13: □□□□□□□□ - □□□□□□□□□□□

DEMOGRAPHICS / Demographics Age

Export Export comparison Show 10 entries Search:

Strata	Patients initiating first-line therapy for hypertension with >1 yr follow-up				Patients initiating first-line therapy for hypertension with >3 yr follow-up				Std diff
	Count	Avg	Std Dev	Median	Count	Avg	Std Dev	Median	
Female	768,180	49.39	9.78	51.00	390,693	49.01	9.03	51.00	-0.0291
All stratas	1,661,604	48.96	10.00	50.00	837,459	48.64	9.26	50.00	-0.0232

Showing 1 to 2 of 2 entries Previous 1 Next

Figure 11.14: □□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□

11.8 R

□□□□R□□□□□

□□□□□□□□□□□□

```
  covariateSettings = covariateSettings,
  aggregated = TRUE)

summary(covariateData2)
```

□□□□□□□□□□□□□□

```
## CovariateData Object Summary
##
## Number of Covariates: 41330
## Number of Non-Zero Covariate Values: 41330
```

11.8.5

□□□□□covariateData□□□□□□□□□□□□□□

11.9.1

□□□1□□□□□3□□□□□□□□□□□□□

Event Cohorts

Each Event Cohort defines the step in a pathway that may occur for a person in the Target Cohort.

Import

Show 10 entries Search:

ID	Name		
9174	ACE inhibitor use	Edit cohort	Remove
9175	Angiotensin receptor blocker (ARB) use	Edit cohort	Remove
9176	Thiazide or thiazide-like diuretic use	Edit cohort	Remove
9177	dihydropyridine Calcium Channel Blocker (dCCB) use	Edit cohort	Remove
9178	non-dihydropyridine Calcium Channel Blocker (ndCCB) use	Edit cohort	Remove
9179	beta blocker use	Edit cohort	Remove
9180	Diuretic-loop use	Edit cohort	Remove
9181	Diuretic-potassium sparing use	Edit cohort	Remove
9182	alpha-1 blocker use	Edit cohort	Remove

Showing 1 to 9 of 9 entries Previous 1 Next

Figure 11.19: □□□□□□

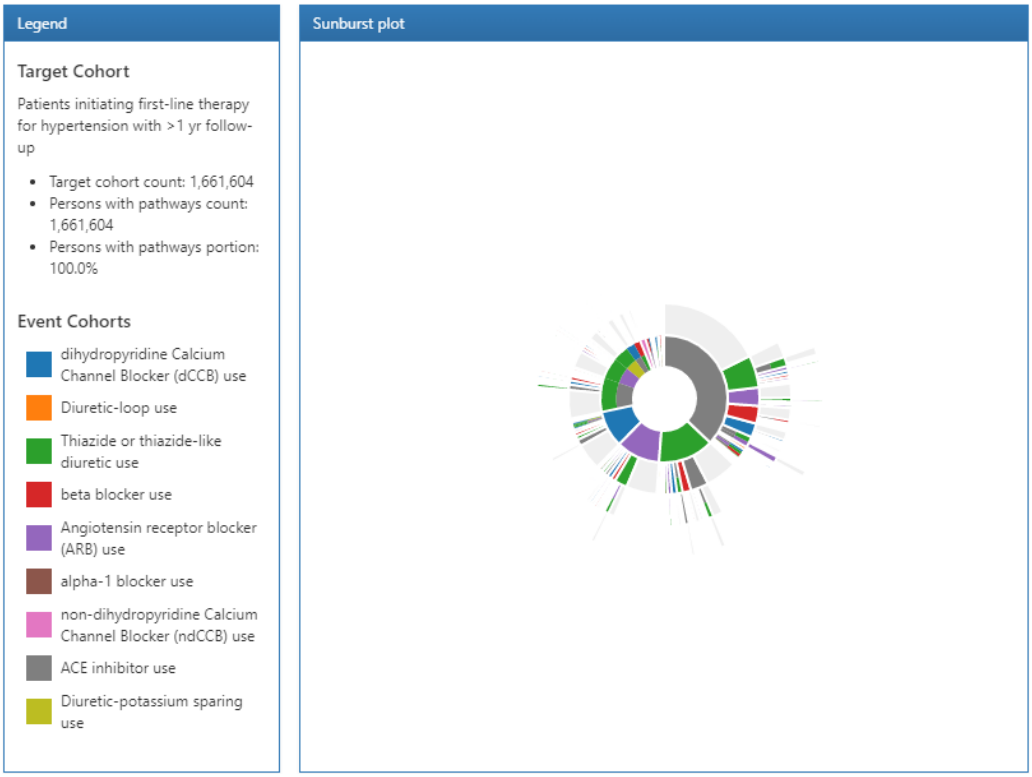

Legend
Target Cohort
Patients initiating first-line therapy for hypertension with >1 yr follow-up
Target cohort count: 1,661,604
Persons with pathways count: 1,661,604
Persons with pathways portion: 100.0%
Event Cohorts
dihydropyridine Calcium Channel Blocker (dCCB) use
Diuretic-loop use
Thiazide or thiazide-like diuretic use
beta blocker use
Angiotensin receptor blocker (ARB) use
alpha-1 blocker use
non-dihydropyridine Calcium Channel Blocker (ndCCB) use
ACE inhibitor use
Diuretic-potassium sparing use
Sunburst plot

Figure 11.20: □□□□□□□□□□□□□□□□□□□□□

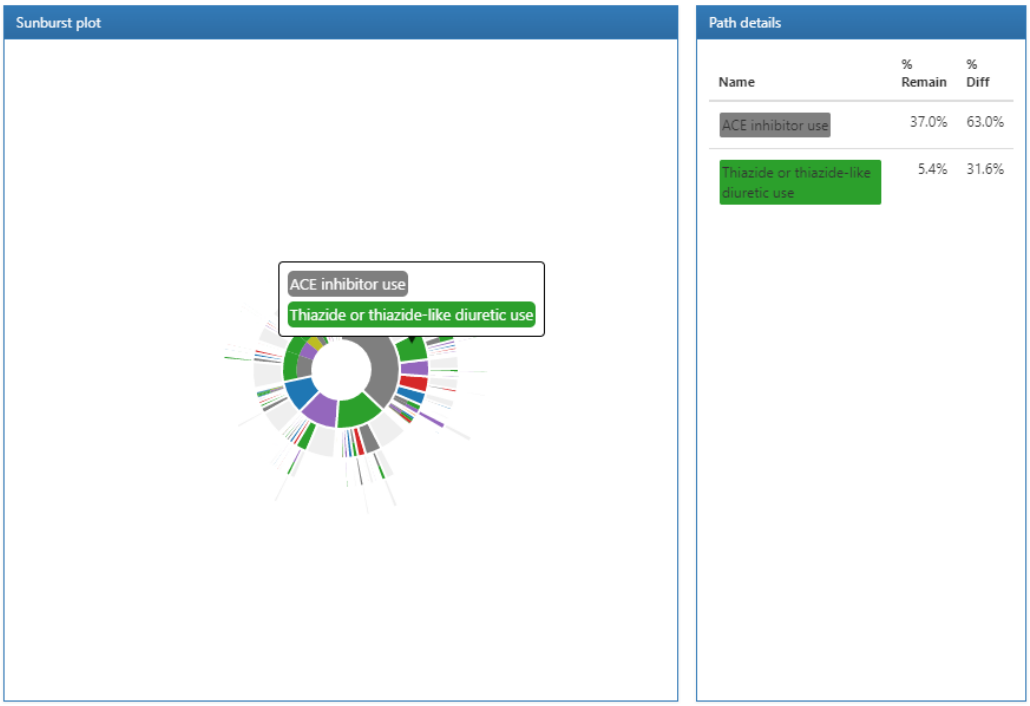


Figure 11.21: □□□□□□□□□□□□□□□□□□□□□□□

11.10 ATLAS

□□□□□□□□□□□□□□□□□□□

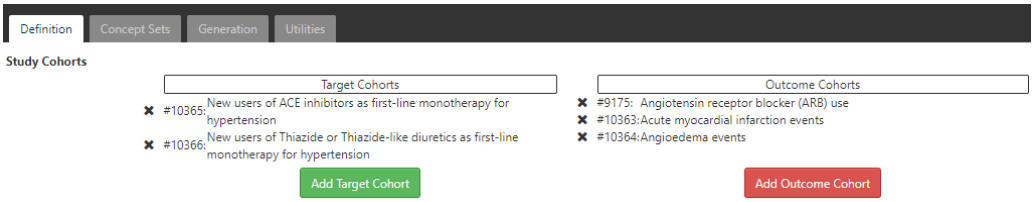


Figure 11.22: □□□□□□□□□□□□□□□□□□□□□

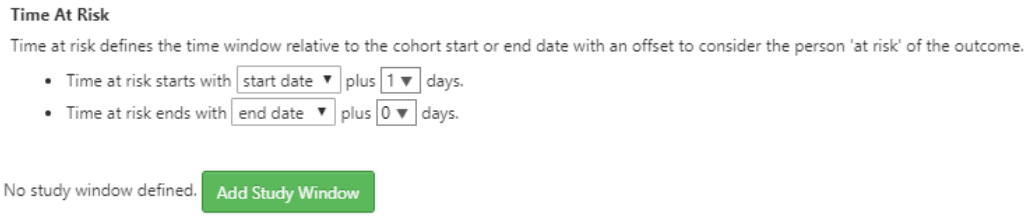


Figure 11.23: □□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□

Figure 11.25: □□□□□□□□

□□□□□□□□□□□□□□□□□ ACEi □□□□□□□□□□□□□□□□□□□□□□ (AMI) □□□□□□□□ Reports □□□□□□□□□□□□□□□□□□□□

ARB □□□□□□□ Reports □□□□□□□□□□□□□□□□□□□□

	Persons	Cases	Proportion [+|-] per 1k persons	Time At Risk (years)	Rate [+|-] per 1k years
Summary Statistics:	818,520	43,868	53.59	621,094	70.63

Stratify Rule	N	Cases	Proportion [+|-] per 1k persons	Time At Risk (years)	Rate [+|-] per 1k years
1. Gender = Female	313,832	21,049	67.07	228,369	92.17

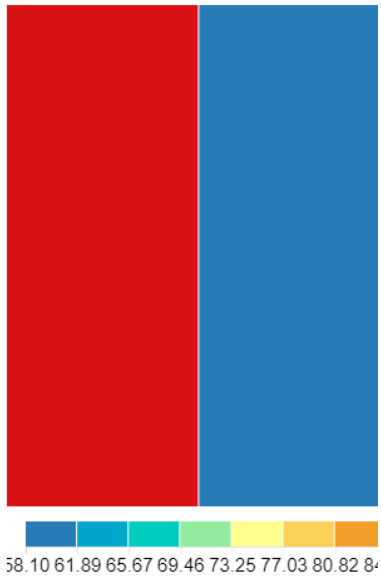


Figure 11.27: □□□ - ACEi □□□□ ARB □□□□□□□□□□□ ACEi □□□

□□□□□□□□□□□□□□□□□□□

□□ **11.2.** □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□365
□□□□□□□□□□□□□□□□□□□

Chapter 12

□□: *Martijn Schuemie, David Madigan, Marc Suchard & Patrick Ryan*

□□□

12.1 □□□□□□□□□□5□□□□□□□□□□□□□□□□□□□□

Table 12.1: □□□□□□□□□□□□□□□□□□□□

□□	□□
□□□□□□□□□□	□□□□□□□□□□□□□□□□□□
□□□□□□□□□	□□□□□□□□□□□□□□□□□
□□□□□□□□□□	□□□□□□□□□□□□□□□□□□□□□□
□□□□□□	□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□

Table 12.2: □□□□□□□□□□□□□□□□□□□□□□□□□

□□	□□
□□□□□□□□□□	□□□□□□□□□□
□□□□□□□□□□	□□□□□□□□□□□□□□□□□□□
□□□□□□	□□□□□□□□□□□□□□□□□□□□

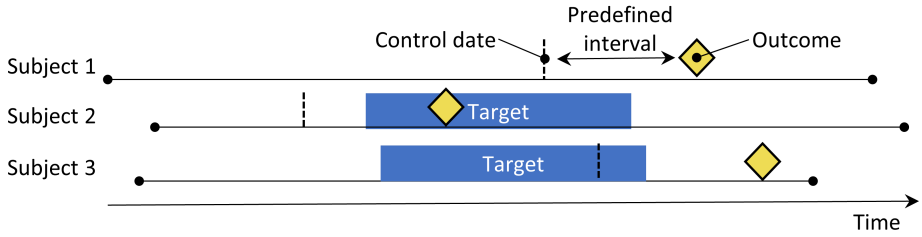

Predefined
interval
Control date
Outcome
Subject 1
Subject 2
Target
Subject 3
Target
Time

Figure 12.4: □□□□□□□□□□□□□□□□

Table 12.5: □□□□□□□□□□□□□□□□□□□□□□□□

□□	□□
□□□□□□□□□□	□□□□□□□□□□□□
□□□□□□□□□□	□□□□□□□□□□□□□□□□□□□□□□
□□□□□□	□□□□□□□□□□□□□□□□□□□□

□□□	□
□□□□□ □□□	□□□□□□□□□□□□□□□□□□□□□□□□□□ □□□□□□□□□□□□□□□□□□□□□□□□□□□□□

12.6.7

□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□

Concept Set #1798551

Concepts to exclude for ACEi and THZ | Optimize

Concept Set Expression | Included Concepts 38225 | Included Source Codes | Explore Evidence | Export | Compare

Show 25 entries | Search:

Showing 1 to 14 of 14 entries | Previous 1 Next

Concept Id	Concept Code	Concept Name	Domain	Standard Concept Caption	Exclude	Descendants	Mapped
1342439	38454	trandolapril	Drug	Standard			
1334456	35296	Ramipril	Drug	Standard			
1331235	35208	quinapril	Drug	Standard			
1373225	54552	Perindopril	Drug	Standard			
1310756	30131	moexipril	Drug	Standard			

Figure 12.8: □□□□□□□□□□□□□□□□□□□□□□□□□□

Negative control concept set:

Negative controls for ACEi and THZ

Covariate selection

Concepts to **include** when constructing the covariates to be used in this study. (Leave blank if you want to include every concept).*

** Concepts defined here are combined with those defined in the Analysis settings section.*

Concepts to **exclude** when constructing the covariates to be used in this study.*

Concepts to exclude for ACEi and THZ

** Concepts defined here are combined with those defined in the Analysis settings section.*

Figure 12.9: □□□□□□□□□□□□□□□□□□□□

**included in the cohort (□□□□□□□□

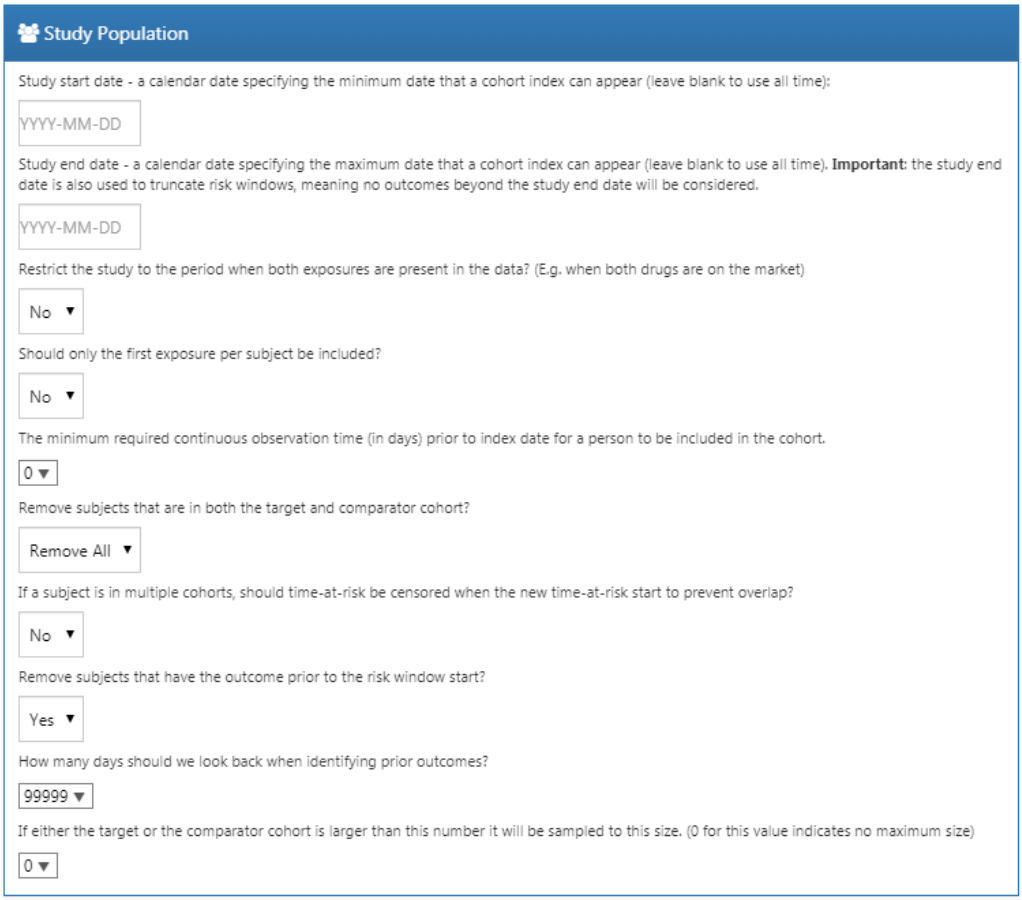


Figure 12.11: □□□□□□□□□□

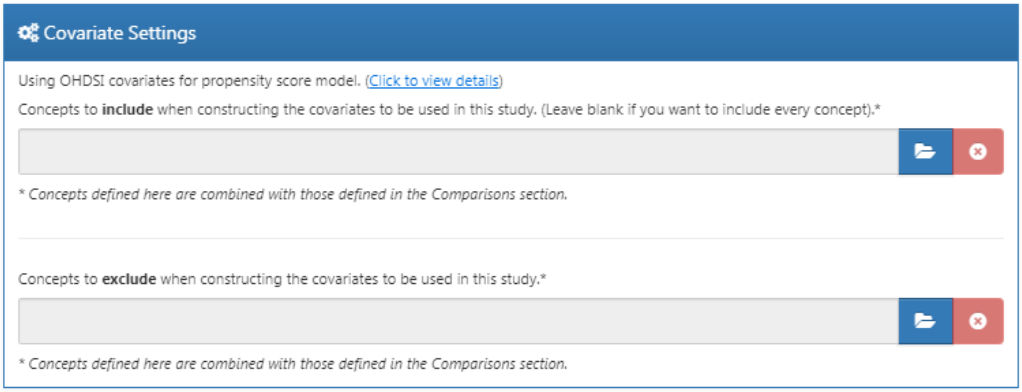


Figure 12.12: □□□□□□

□□□□□□□

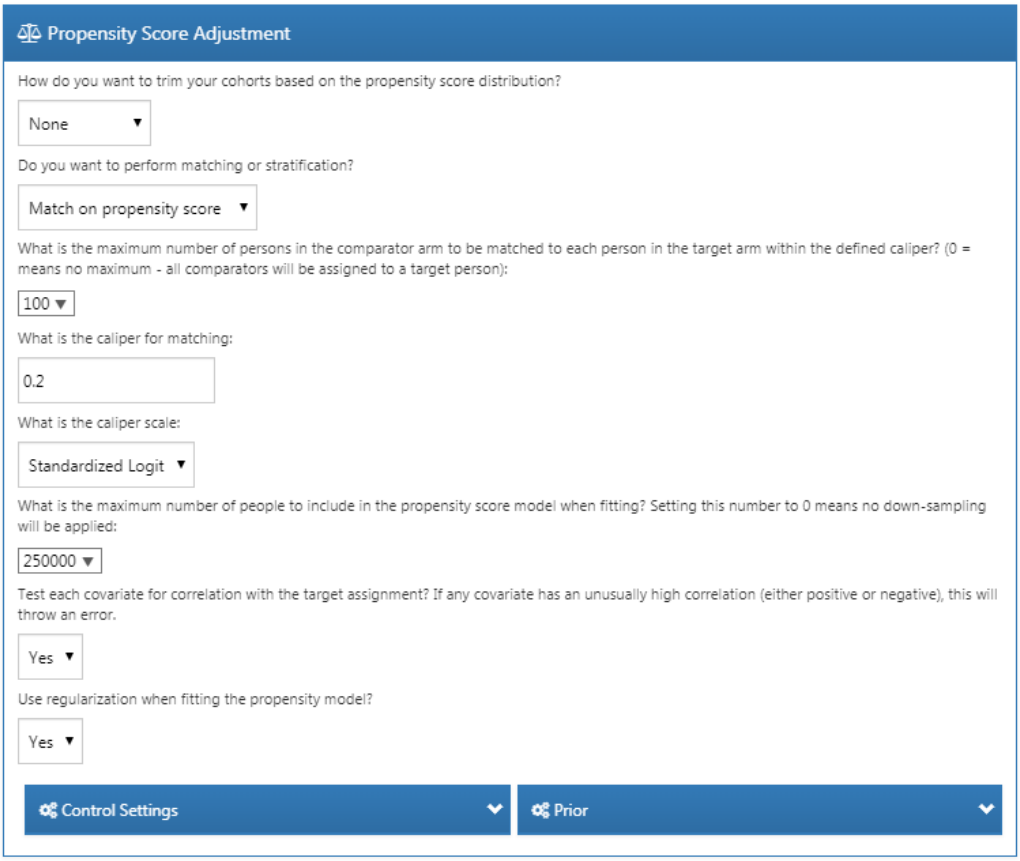

Propensity Score Adjustment
How do you want to trim your cohorts based on the propensity score distribution?
None
Do you want to perform matching or stratification?
Match on propensity score
What is the maximum number of persons in the comparator arm to be matched to each person in the target arm within the defined caliper? (0 = means no maximum - all comparators will be assigned to a target person):
100
What is the caliper for matching:
0.2
What is the caliper scale:
Standardized Logit
What is the maximum number of people to include in the propensity score model when fitting? Setting this number to 0 means no down-sampling will be applied:
250000
Test each covariate for correlation with the target assignment? If any covariate has an unusually high correlation (either positive or negative), this will throw an error.
Yes
Use regularization when fitting the propensity model?
Yes
Control Settings
Prior

Figure 12.14: □□□□□□□□□□□□□□□

□□□□□□□□

Negative Control Outcome Cohort Definition

This expression will define the criteria for inclusion and duration of time for cohorts intended for use as negative control outcomes.
The type of occurrence of the event when selecting from the domain.

First occurrence

When true, descendant concepts for the negative control outcome concept IDs will be used to detect the outcome and roll up the occurrence to the concept ID.

Yes

What domains should be considered to detect negative control outcomes? (Hold control to select multiple domains)

Condition
Drug
Device
Measurement
Observation
Procedure
Visit

Figure 12.16: □□□□□□□□□□□□□□□□□□□□□□□□□□□□

Positive Control Synthesis

Should we perform positive control synthesis? (to calibrate confidence intervals)

Yes

Model Type:

Survival

Using OHDSI covariates for model. (Click to view details)

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

1 days from cohort start date

Define the time-at-risk window end:

0 days from cohort end date

The minimum required continuous observation time (in days) prior to exposure:

365

Should only the first exposure per subject be included?

Yes

Should only the first outcome per person be considered when modeling the outcome?

Yes

Remove people with prior outcomes?

Yes

Advanced Settings start here

Additional Settings

Figure 12.17: □□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□“Download (□□□□□□)” □□□□□□□□zip□□□□□□□□□□□□□□□□□zip□□□
(Wickham, 2015)□□□□□□□□□□□□□□□□□□R Studio □□□□□□□□□□□□R Studio
□□□□□□□□□□□□□□□□□□□□□□□□□□□□.Rproj □□□□□□□□□□□□□□□□ R
Studio □□□□□□□R□□□□□□R□□□□□□□□□□□□□□□□□□□□□□□ Upload □□□□□□□□□□□□□□□
R Studio□□□□□□□□□□□□□□□README□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“High correlation between co-variate(s) and treatment detected.”□□□□□□□□□□□□□□□□□□□□□□□□□□□”
□□□□□□□□□□□□□□□□□□□

```
#
aceI <- c(1335471,1340128,1341927,1363749,1308216,1310756,1373225,
          1331235,1334456,1342439)
thz <- c(1395058,974166,978555,907013)

#
cs <- createDefaultCovariateSettings(excludedCovariateConceptIds = c(aceI,
                                                                     thz),
                                     addDescendantsToExclude = TRUE)

#
cmData <- getDbCohortMethodData(connectionDetails = connectionDetails,
                                cdmDatabaseSchema = cdmDatabaseSchema,
                                oracleTempSchema = NULL,
                                targetId = 1,
                                comparatorId = 2,
                                outcomeIds = 3,
                                studyStartDate = "",
                                studyEndDate = "",
                                exposureDatabaseSchema = cohortDbSchema,
                                exposureTable = cohortTable,
                                outcomeDatabaseSchema = cohortDbSchema,
                                outcomeTable = cohortTable,
                                cdmVersion = cdmVersion,
                                firstExposureOnly = FALSE,
                                removeDuplicateSubjects = FALSE,
                                restrictToCommonPeriod = FALSE,
                                washoutPeriod = 0,
                                covariateSettings = cs)
cmData
```

```
## CohortMethodData
##
##     ID 1
##      ID 2
##       ID(s) 3
```

□□□□□□□□□□□□□□□□□□□□□CohortMethod□□□□□□□□□□□□□□□□□□□createDefault
(12.1)□□□□□□□□□□□□□□□□□□□

```
##
##     ID 1
##      ID 2
##       ID(s) 3
##
##      67166
##     35333
##
##
##
## 3                980          891
##
##
##    58349
##        24484665
```

cohortMethodData□□□□□□□□

```
                                    riskWindowEnd = 0,
                                    endAnchor = "cohort end")
```

firstExposureOnly□removeDuplicateSubjects□FALSE□□□□□washoutPeriod□0□□□
= 1□□□startAnchor = "cohort start"□□□□□□□□□□□

```
##
##                95%    95%   logRr    seLogRr
##        4.3203   2.4531   8.0771   1.4633   0.304
```

12.8.6

□□□□□□□□

```
  firstExposureOnly = FALSE,
  restrictToCommonPeriod = FALSE,
  washoutPeriod = 0,
  removeDuplicateSubjects = "remove all",
  removeSubjectsWithPriorOutcome = TRUE,
  minDaysAtRisk = 1,
  startAnchor = "cohort start",
  addExposureDaysToStart = FALSE,
  endAnchor = "cohort end",
  addExposureDaysToEnd = TRUE)

psArgs <- createCreatePsArgs()

matchArgs <- createMatchOnPsArgs(
  caliper = 0.2,
  caliperScale = "standardized logit",
  maxRatio = 100)

fomArgs <- createFitOutcomeModelArgs(
  modelType = "cox",
  stratified = TRUE)
```

□□□□□□□1□□□□□□□□□□□□□□□□□□□□□□□□□□ID□□□□□□□□□□□□□□□□□1□□□□□

```
cmAnalysis <- createCmAnalysis(
  analysisId = 1,
  description = "Propensity score matching",
  getDbCohortMethodDataArgs = cmdArgs,
  createStudyPopArgs = spArgs,
  createPs = TRUE,
  createPsArgs = psArgs,
  matchOnPs = TRUE,
  matchOnPsArgs = matchArgs
  fitOutcomeModel = TRUE,
  fitOutcomeModelArgs = fomArgs)

cmAnalysisList <- list(cmAnalysis)
```

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

```
result <- runCmAnalyses(connectionDetails = connectionDetails,
                        cdmDatabaseSchema = cdmDatabaseSchema,
                        exposureDatabaseSchema = cohortDbSchema,
                        exposureTable = cohortTable,
                        outcomeDatabaseSchema = cohortDbSchema,
                        outcomeTable = cohortTable,
                        cdmVersion = cdmVersion,
```

```
                        outputFolder = outputFolder,
                        cmAnalysisList = cmAnalysisList,
                        targetComparatorOutcomesList = tcosList)
```

result□□□AMI□□□□□□□□□□□□□□

```
omFile <- result$outcomeModelFile[result$targetId == 1 &
                                  result$comparatorId == 2 &
                                  result$outcomeId == 4 &
                                  result$analysisId == 1]
outcomeModel <- readRDS(file.path(outputFolder, omFile))
outcomeModel
```

```
## Model type: cox
## Stratified: TRUE
## Use covariates: FALSE
## Use inverse probability of treatment weighting: FALSE
## Status: OK
##
##                95%   95%  logRr   seLogRr
##     1.1338    0.5921    2.1765 0.1256   0.332
```

□□□1□□□□□□□□□□□□□□□□□□□

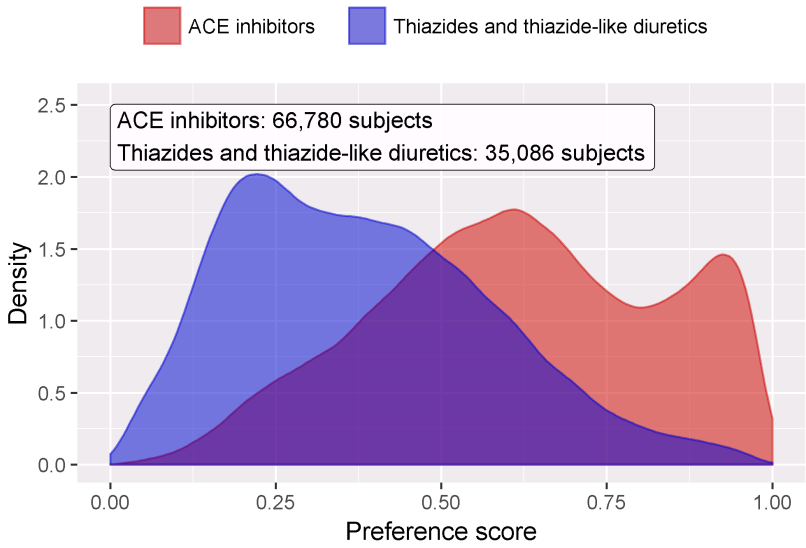

ACE inhibitors
Thiazides and thiazide-like diuretics
ACE inhibitors: 66,780 subjects
Thiazides and thiazide-like diuretics: 35,086 subjects
Density
2.5
2.0
1.5
1.0
0.5
0.0
0.00
0.25
0.50
0.75
1.00
Preference score

Figure 12.18: □□□□□□□□

□□□□□□□

12.9.2

PS□□□□□□□□□□2□□□□□□□□□□

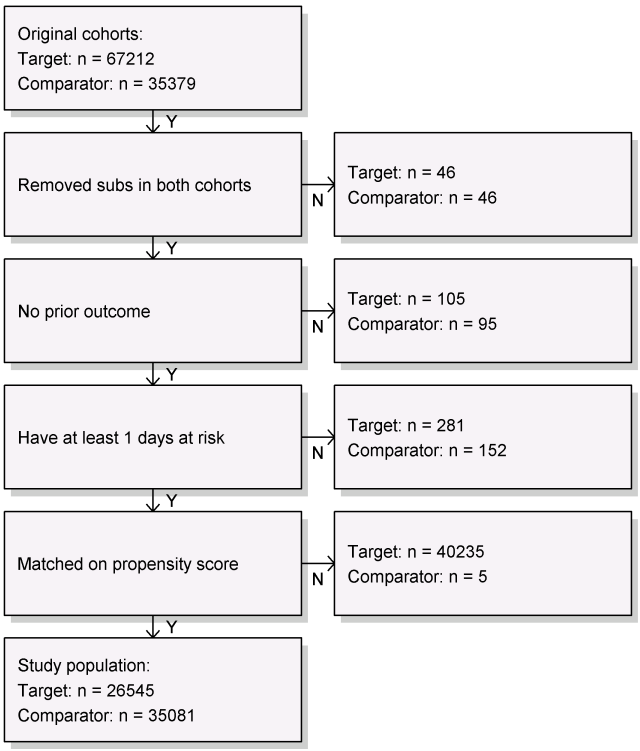

Original cohorts:
Target: n = 67212
Comparator: n = 35379
Y
Removed subs in both cohorts
N
Target: n = 46
Comparator: n = 46
Y
No prior outcome
N
Target: n = 105
Comparator: n = 95
Y
Have at least 1 days at risk
N
Target: n = 281
Comparator: n = 152
Y
Matched on propensity score
N
Target: n = 40235
Comparator: n = 5
Y
Study population:
Target: n = 26545
Comparator: n = 35081

Figure 12.20: □□□□□□

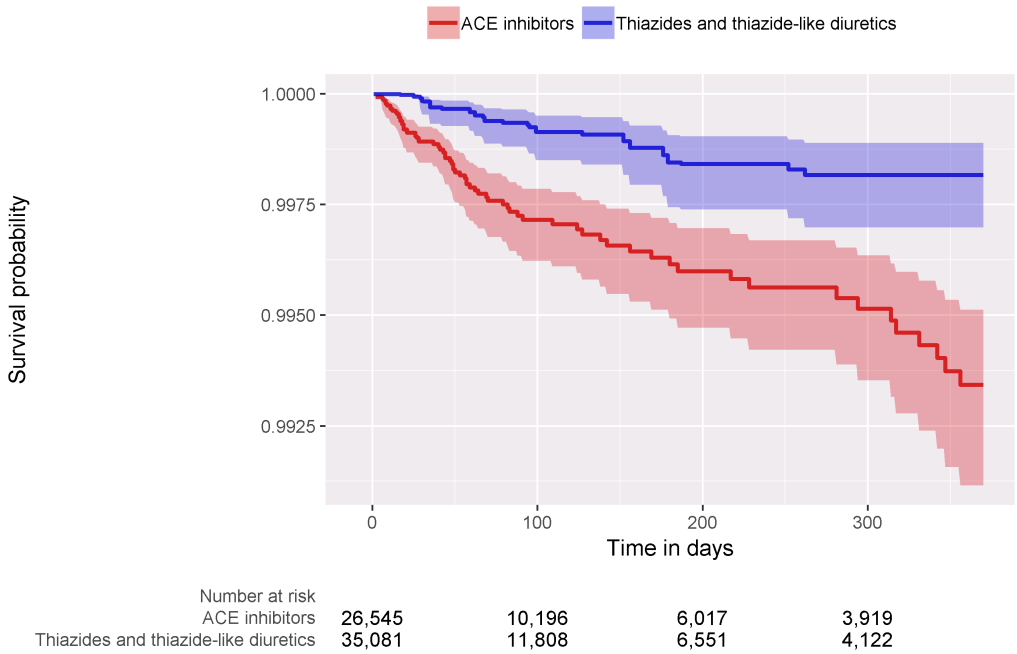


Figure 12.22: □□□□–□□□□□□□□

12.9.5

□□□□□□□□□□□□□□□□□□□□4.32□95□□□□□□□2.45 - 8.08□□□□□□□□□□□□□□□ACEi□THZ□ - 2.18□□□□□□□AMI□□□□□□□□□□□□□□□□□□□

```
install.packages(c("SqlRender", "DatabaseConnector", "remotes"))
remotes::install_github("ohdsi/Eunomia", ref = "v1.0.0")
remotes::install_github("ohdsi/CohortMethod")
```

Eunomia□□□□□□□□□□□□□□R□□□□□□□□□□□□□□CDM□□□□□□□□□□□□□□

Chapter 13

□□: *Peter Rijnbeek & Jenna Reps*

□□□

□□□	□□
□□□□□□□□□□	□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□	t=0□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□	□□□□□□□□□□□□□□□□□□□□

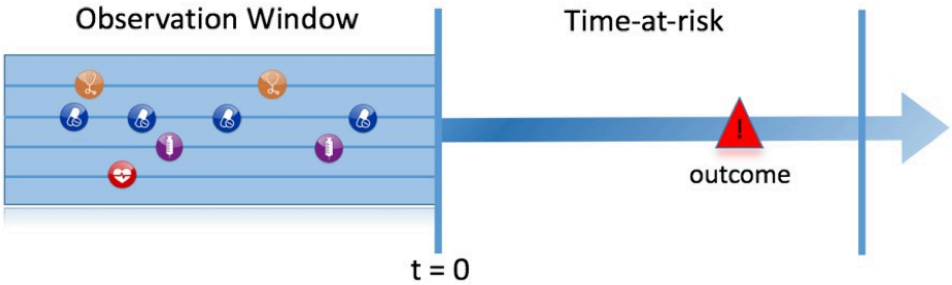


Figure 13.1: □□□□□

□ 13.2.1 □□□□□□CONDITION_OCCURRENCE□□□□□□□□□□□□□Concept ID 320128 □□□□□□□□□□□□□□□□□□

□□ □□□□CONDITION_OCCURRENCE□□□□□□□□□□□□3□□□□□□□□□□□□□□

PERSON_ID	CONDITION_CONCEPT_ID	CONDITION_START_DATE
1	320128	2000-10-01
2	320128	2001-05-01

□□□□□□□□□□□□□□□□□

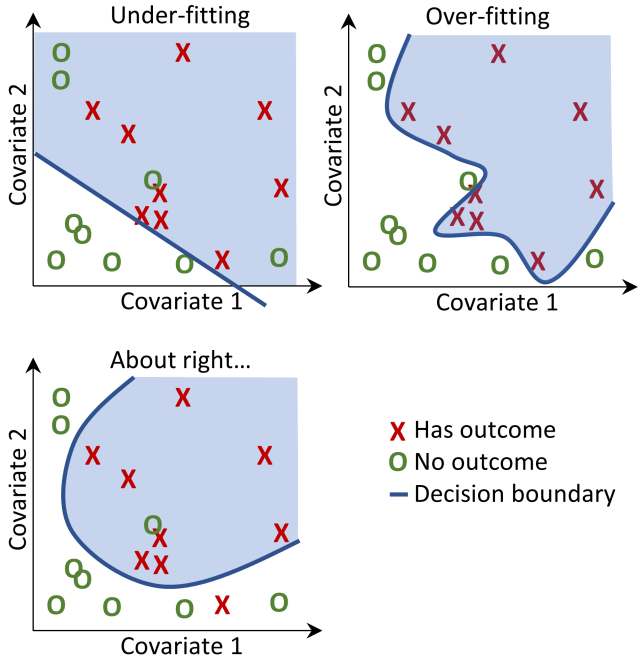


Figure 13.2: Decision boundary.

□□□□□□□

Table 13.6: □□□□□□□□□□□□□□□□□□□□□

□□□□□	□□	□□□□□
maxDepth	□□□□□□	4,10,17
mtries	□□□□□□	-1 = □□□□□□□□□,5,20
ntrees	□□□	500

13.3.4 K-

K-nearest neighbors (KNN) □□□□□□□□□□□□□□□□□□□□

□□□□□	□□	□□□□□
minImpuritySplit	□□□□□□□

  - □□□□□□□□□

13.17 □□□□□□□□□□□□□□□□□□□□□□□□□□AUC□□□□□□□□□□□□□□□□□□□□□□□□0.5□□□□

AUC□□□□□□□□□□□□□□□□□□□

13.5.1

□□□□□□□□□ACE□□□□□□□□□□□□□□□□□□□□□ACE□□□□□□□□□□□□□□□□□□□

□□	□
□□□□□□□□□□ □□□□□ □□□	□□□□□□□□ □□□□□□□□1□□□365□□□□□□□□364□□□□□□□□□□ Gradient Boosting Machine with hyper-parameters ntree: 5000, max depth: 4 or 7 or 10 and learning rate: 0.001 or 0.01 or 0.1 or 0.9. Covariates will include gender, age, conditions, drugs, drug groups, and visit count. □□□□□: 75□□□□□□□□ - 25□□□□□□□□□□□□□□□□□□□□□□□□□□□□

13.6 ATLAS

□□□□□□□□□□□□□□□□□□□□□□ATLAS□□□□□□□□□□□□□ Prediction □□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□

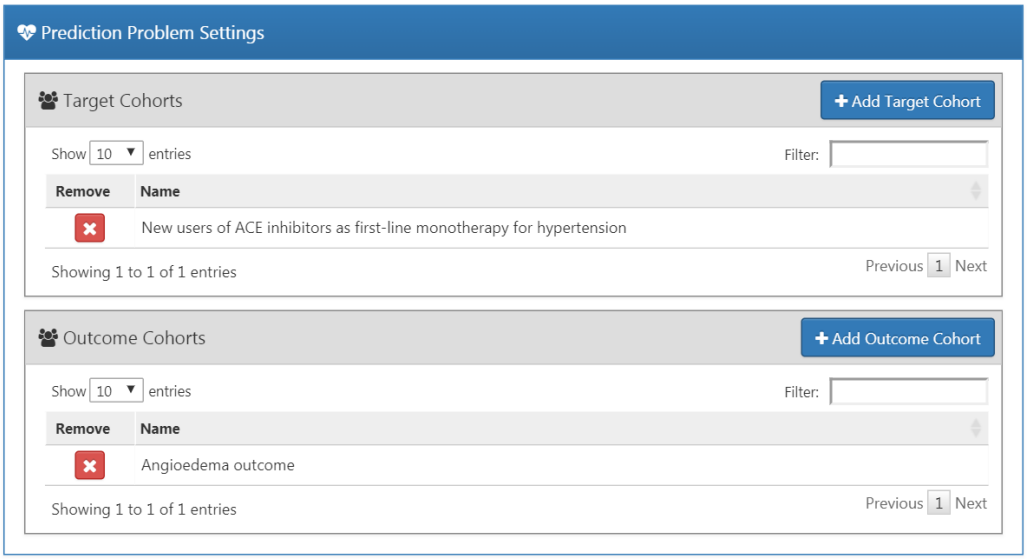

Prediction Problem Settings
Target Cohorts
Add Target Cohort
Show 10 entries
Filter:
Remove
Name
New users of ACE inhibitors as first-line monotherapy for hypertension
Showing 1 to 1 of 1 entries
Previous 1 Next
Outcome Cohorts
Add Outcome Cohort
Show 10 entries
Filter:
Remove
Name
Angioedema outcome
Showing 1 to 1 of 1 entries
Previous 1 Next

Figure 13.4: □□□□□□.

□□□□□□□□□□□□□□□□□□/□□□□□□□□□□□□□□□□□□

Gradient Boosting Machine Model Settings
Use the options below to edit the model settings

The boosting learn rate (default = 0.01,0.1):

Boosting learn rate	Action
0.001	Remove
0.01	Remove
0.1	Remove
0.9	Remove
	Add / Reset to default

Maximum number of interactions - a large value will lead to slow model training (default = 4,6,17):

Maximum number of interactions	Action
4	Remove
7	Remove
10	Remove
	Add / Reset to default

The minimum number of rows required at each end node of the tree (default = 20):

Minimum number of rows	Action
20	Remove
	Add / Using default

The number of trees to build (default = 10,100):

Trees to build	Action
5000	Remove
	Add / Reset to default

The number of computer threads to use (how many cores do you have?) (default = 20):

20 Using default

Figure 13.5: □□□□□□□□□□□□□□□□

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

Should descendant concepts be added to the list of included concepts?

No

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

Should descendant concepts be added to the list of included concepts?

No

A comma delimited list of covariate IDs that should be restricted to:

Figure 13.6: □□□□□□□□□□□□.

- □□□□□□□[□□□□□□□□□□□□□]□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

- □□□□□□□□ □□□□□□□□□□□□□□□□□□□□□□ID□□□□□□□□□□□□□□□□□□□□CONDITON_
  1□□□□□□□□□□ 0□
- □□□□□□□□□□□□□ □□□□□□□□□□□□□□□□□□□□□□ID□□□□□□□□□□□□□□□□□□□□CONDI
  1□□□□□□□□□□ 0□
- □□□□□□□□□□□□□□□□□□□□ID□□□□□□□□□□□□□□□□□□□□DRUG_ERA□□□□□□□□□□□□
  1□□□□□□□□□□ 0□
- □□□□□□□□□□□□□□□□□□□□□□□□ID□□□□□□□□□□□□□□□□□□□□DRUG_ERA□□□□□□□□
  1□□□□□□□□□□ 0□

□□□□□□□□□□□□□□□□□□□

- □□□□□□□□□

- Visit Count: Construct covariates for each visit and time interval selected and count the number of visits recorded during the time interval as the covariate value.
- Visit Concept Count: Construct covariates for each visit, domain and time interval selected and count the number of records per domain recorded during the visit type and time interval as the covariate value.

The distinct count option counts the number of distinct concept IDs per domain and time interval.

Once done, this section should look like Figure 13.10.

Set the time bound covariates

Domain	Any Time Prior	Long Term (-365 days)	Medium Term (-180 days)	Short Term (-30 days)	Distinct Count		
					Long Term (-365 days)	Medium Term (-180 days)	Short Term (-30 days)
Condition	☐	☐	☐	☐	☐	☐	☐
Condition - Primary Inpatient	☐	☐	☐	☐			
Drug	☐	☐	☐	☐	☐	☐	☐
Procedure	☐	☑	☐	☐	☐	☐	☐
Measurement	☐	☑	☐	☐	☐	☐	☐
Measurement - Value	☐	☐	☐	☐			
Measurement - Range Group	☐	☑	☐	☐			
Observation	☐	☑	☐	☐	☐	☐	☐
Device	☐	☑	☐	☐			
Visit - Count		☑	☐	☐			
Visit - Concept Count		☐	☐	☐			

Figure 13.10: Time bound covariates.

The final option is whether to include commonly used risk scores as covariates. Once done, the risk score settings should look like Figure 13.11.

Set the index score covariates

Index Score Type	
$CHADS_2$	☐
CHA_2DS_2VASc	☑
DCSI	☑
Charlson	☑

Figure 13.11: Risk score covariate settings.

Population Settings

The population settings is where addition inclusion criteria can be applied to the target population and is also where the time-at-risk is defined. To add a population setting into the study, click on the "Add Population Settings" button. This will open up the population setting view.

The first set of options enable the user to specify the time-at-risk period. This is the time interval where we look to see whether the outcome of interest occurs. If a patient has the outcome during the time-at-risk period then we will classify them as "Has outcome", otherwise they are classified as "No outcome". "**Define the time-at-risk window start, relative to target cohort entry:**" defines the start of the time-at-risk, relative to the target cohort start or end date. Similarly, "**Define the time-at-risk window end:**" defines the end of the time-at-risk.

"**Minimum lookback period applied to target cohort**" specifies the minimum baseline period, the minimum number of days prior to the cohort start date that a patient is continuously observed. The default is 365 days. Expanding the minimum look-back will give a more complete picture of a patient (as they must have been observed for longer) but will filter patients who do not have the minimum number of days prior observation.

If "**Should subjects without time at risk be removed?**" is set to yes, then a value for "**Minimum time at risk:**" is also required. This allows removing people who are lost to follow-up (i.e. that have left the database during the time-at-risk period). For example, if the time-at-risk period was 1 day from cohort start until 365 days from cohort start, then the full time-at-risk interval is 364 days (365-1). If we only want to include patients who are observed the whole interval, then we set the minimum time at risk to be 364. If we are happy as long as people are in the time-at-risk for the first 100 days, then we select minimum time at risk to be 100. In this case as the time-at-risk start is 1 day from the cohort start, a patient will be included if they remain in the database for at least 101 days from the cohort start date. If we set "Should subjects without time at risk be removed?" to 'No', then this will keep all patients, even those who drop out from the database during the time-at-risk.

The option "**Include people with outcomes who are not observed for the whole at risk period?**" is related to the previous option. If set to "yes", then people who experience the outcome during the time-at-risk are always kept, even if they are not observed for the specified minimum amount of time.

The option "**Should only the first exposure per subject be included?**" is only useful if our target cohort contains patients multiple times with different cohort start dates. In this situation, picking "yes" will result in only keeping the earliest target cohort date per patient in the analysis. Otherwise a patient can be in the dataset multiple times.

Setting "**Remove patients who have observed the outcome prior to cohort entry?**" to "yes" will remove patients who have the outcome prior to the time-at-risk start date, so the model is for patients who have never experienced the outcome before. If "no" is selected, then patients could have had the outcome prior. Often, having the outcome prior is very predictive of having the outcome during the time-at-risk.

Once done, the population settings dialog should look like Figure 13.12.

Now that we are finished with the Analysis Settings, the entire dialog should look like Figure 13.13.

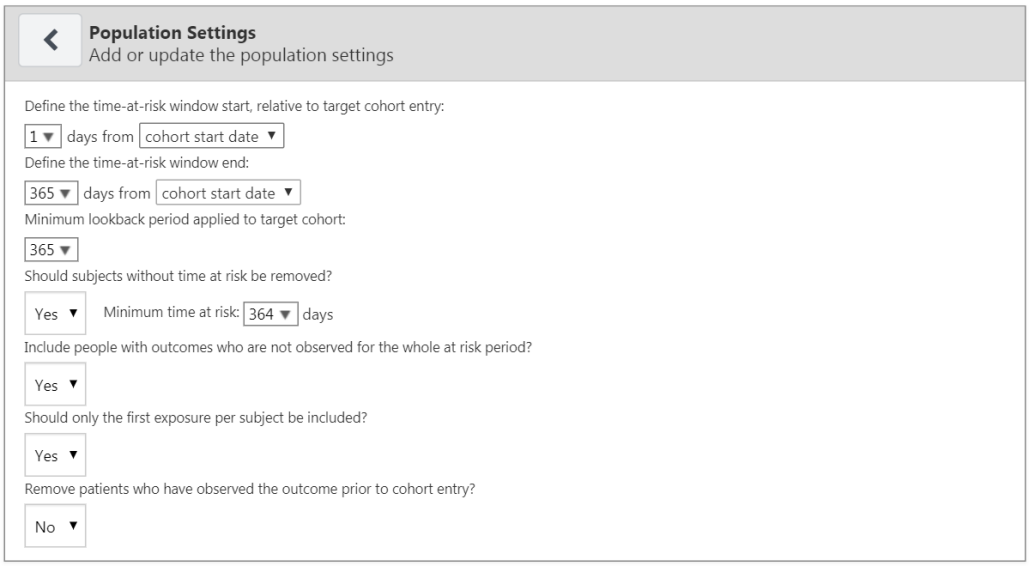


Figure 13.12: Population settings.

13.6.3 Execution Settings

There are three options:

- "**Perform sampling**": here we choose whether to perform sampling (default = "no"). If set to "yes", another option will appear: "**How many patients to use for a subset?**", where the sample size can be specified. Sampling can be an efficient means to determine if a model for a large population (e.g. 10 million patients) will be predictive by creating and testing the model with a sample of patients. For example, if the AUC is close to 0.5 in the sample, we might abandon the model.
- "**Minimum covariate occurrence: If a covariate occurs in a fraction of the target population less than this value, it will be removed:**": here we choose the minimum covariate occurrence (default = 0.001). A minimum threshold value for covariate occurrence is necessary to remove rare events that are not representative of the overall population.
- "**Normalize covariate**": here we choose whether to normalize covariates (default = "yes"). Normalization of the covariates is usually necessary for successful implementation of a LASSO model.

For our example we make the choices shown in Figure 13.14.

13.6.4 Training Settings

There are four options:

- "**Specify how to split the test/train set:**" Select whether to differentiate the train/test data by person (stratified by outcome) or by time (older data to train the model, later data to evaluate the model).

Analysis Settings

Model Settings

+ Add Model Settings

Show 10 entries Filter:

Remove	Model	Options
✖	GradientBoostingMachineSettings	{"ntrees":[5000],"nthread":20,"maxDepth":[4,7,10],"minRows":[20],"learnRate":[0.001,0.01,0.1,0.9],"seed":null}

Showing 1 to 1 of 1 entries Previous 1 Next

Covariate Settings

+ Add Covariate Settings

Column visibility Copy CSV Show 10 entries Filter:

Remove	Options
✖	DemographicsGender, DemographicsAgeGroup, DemographicsRace, DemographicsEthnicity, DemographicsIndexMonth, ConditionGroupEraLongTerm (+12 more covariate settings)

Showing 1 to 1 of 1 entries Previous 1 Next

Population Settings

+ Add Population Settings

Column visibility Copy CSV Show 10 entries Filter:

Remove	Risk Window Start	Risk Window End	Washout Period	Include All Outcomes	Remove Subjects With Prior Outcome	Minimum Time At Risk
✖	1d from cohort start date	365d from cohort start date	365d	true	false	364d

Showing 1 to 1 of 1 entries Previous 1 Next

Figure 13.13: Analysis settings.

Execution Settings

Perform sampling:

Yes

How many patients to use for a subset? 500000 patients

Minimum covariate occurrence: If a covariate occurs in a fraction of the target population less than this value, it will be removed:

0.001

Normalize covariates:

Yes

Figure 13.14: Execution settings.

- "**Percentage of the data to be used as the test set (0-100%)**": Select the percentage of data to be used as test data (default = 25%).
- "**The number of folds used in the cross validation**": Select the number of folds for cross-validation used to select the optimal hyper-parameter (default = 3).
- "**The seed used to split the test/train set when using a person type testSplit (optional):**": Select the random seed used to split the train/test set when using a person type test split.

For our example we make the choices shown in Figure 13.15.

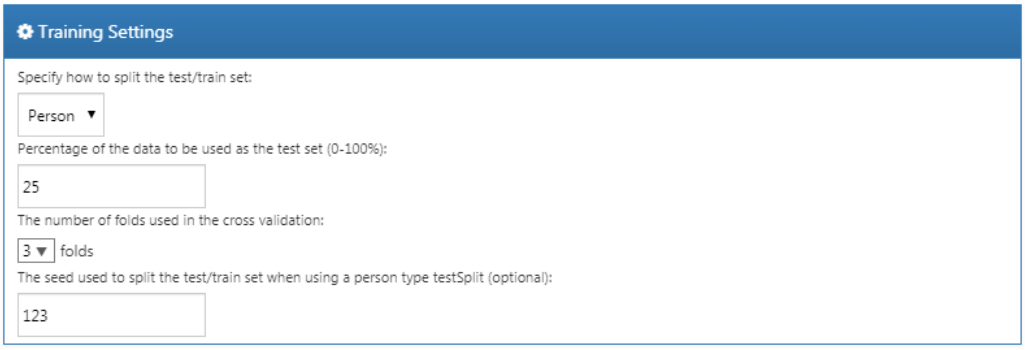


Figure 13.15: Training settings.

13.6.5 Importing and Exporting a Study

To export a study, click on the "Export" tab under "Utilities." ATLAS will produce JSON that can be directly copied and pasted into a file that contains all of the data, such as the study name, cohort definitions, models selected, covariates, settings, needed to run the study.

To import a study, click on the "Import" tab under "Utilities." Paste the contents of a patient-level prediction study JSON into this window, then click on the Import button below the other tab buttons. Note that this will overwrite all previous settings for that study, so this is typically done using a new, empty study design.

13.6.6 Downloading the Study Package

Click on the "Review & Download" tab under "Utilities." In the "Download Study Package" section, enter a descriptive name for the R package, noting that any illegal characters in R will automatically be removed from the file name by ATLAS. Click on

to download the R package to a local folder.

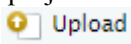
13.6.7 Running the Study

To run the R package requires having R, RStudio, and Java installed as described in Section 8.4.5. Also required is the PatientLevelPrediction package, which can be installed in

R using:

```
install.packages("drat")
drat::addRepo("OHDSI")
install.packages("PatientLevelPrediction")
```

Some of the machine learning algorithms require additional software to be installed. For a full description of how to install the PatientLevelPrediction package, see the "Patient-Level Prediction Installation Guide" vignette.

To use the study R package we recommend using R Studio. If you are running R Studio locally, unzip the file generated by ATLAS, and double click the .Rproj file to open it in R Studio. If you are running R Studio on an R studio server, click 

to upload and unzip the file, then click on the .Rproj file to open the project.

Once you have opened the project in R Studio, you can open the README file, and follow the instructions. Make sure to change all file paths to existing paths on your system.

13.7 R

□□□□□□□ATLAS□□□□□□□□□□□R□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Pa
CDM□□□□□□□□□□□□□□□□□□□

<- "my_cdm_data.dbo"□□□□□□□□□□□

□□□□□□□□□□□

```
savePlpData(plpData, "angio_in_ace_data")
```

□□□□□□□□□□□□□□□□□□□□□□□loadPlpData()□□□□□□□□□□

13.7.3

□□□□□□□□□□□□□□□□2□□□□□□□□□□□□□□□□□□□

```
                    modelSettings = gbmModel,
                    testSplit = 'person',
                    testFraction = 0.25,
                    nfold = 2,
                    splitSeed = 1234)
```

□□□□□□□□□□□□□R□xgboost□□□□□□□□□□□□75%□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

runPlp□□□□□□plpData□plpResults□plpPlots□evaluation□□□□□□□□□□□□□□□□□□□

```
#
plpModel <- loadPlpModel("model")

#   plpData
plpData <- loadPlpData("newData")

population <- createStudyPopulation(plpData = plpData,
                                    outcomeId = 2,
                                    washoutPeriod = 364,
                                    firstExposureOnly = FALSE,
                                    removeSubjectsWithPriorOutcome = TRUE,
                                    priorOutcomeLookback = 9999,
                                    riskWindowStart = 1,
                                    riskWindowEnd = 365,
                                    addExposureDaysToStart = FALSE,
                                    addExposureDaysToEnd = FALSE,
                                    minTimeAtRisk = 364,
                                    requireTimeAtRisk = TRUE,
                                    includeAllOutcomes = TRUE
)

# apply the trained model on the new data
validationResults <- applyModel(population, plpData, plpModel)
```

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□externalValidatePlp □□□□□□□□□□□□result <- runPlp(...) □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□ ID 1 □ 2 □□□□□ mainschema.dob.cohort □□□□CDM □□□□□□□□□ cdmschema.dob □□□□□□□□□□□

```
valResult <- externalValidatePlp(
    plpResult = result,
    connectionDetails = connectionDetails,
    validationSchemaTarget = 'mainschema.dob',
    validationSchemaOutcome = 'mainschema.dob',
    validationSchemaCdm = 'cdmschema.dbo',
    databaseNames = 'new database',
    validationTableTarget = 'cohort',
    validationTableOutcome = 'cohort',
    validationIdTarget = 1,
    validationIdOutcome = 2
)
```

□□□□□□□□□□□□□□□□□□□

```
    validationSchemaTarget = list('mainschema.dob',
                                  'difschema.dob',
                                  'anotherschema.dob'),
    validationSchemaOutcome = list('mainschema.dob',
                                   'difschema.dob',
                                   'anotherschema.dob'),
    validationSchemaCdm = list('cdms1chema.dbo',
                               'cdm2schema.dbo',
                               'cdm3schema.dbo'),
    databaseNames = list('new database 1',
                         'new database 2',
                         'new database 3'),
    validationTableTarget = list('cohort1',
                                 'cohort2',
                                 'cohort3'),
    validationTableOutcome = list('cohort1',
                                  'cohort2',
                                  'cohort3'),
    validationIdTarget = list(1,3,5),
    validationIdOutcome = list(2,4,6)
)
```

13.8

13.8.1

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□viewPlp□□□□□□□□□□□□□□□□□□□□□□□□□□□□

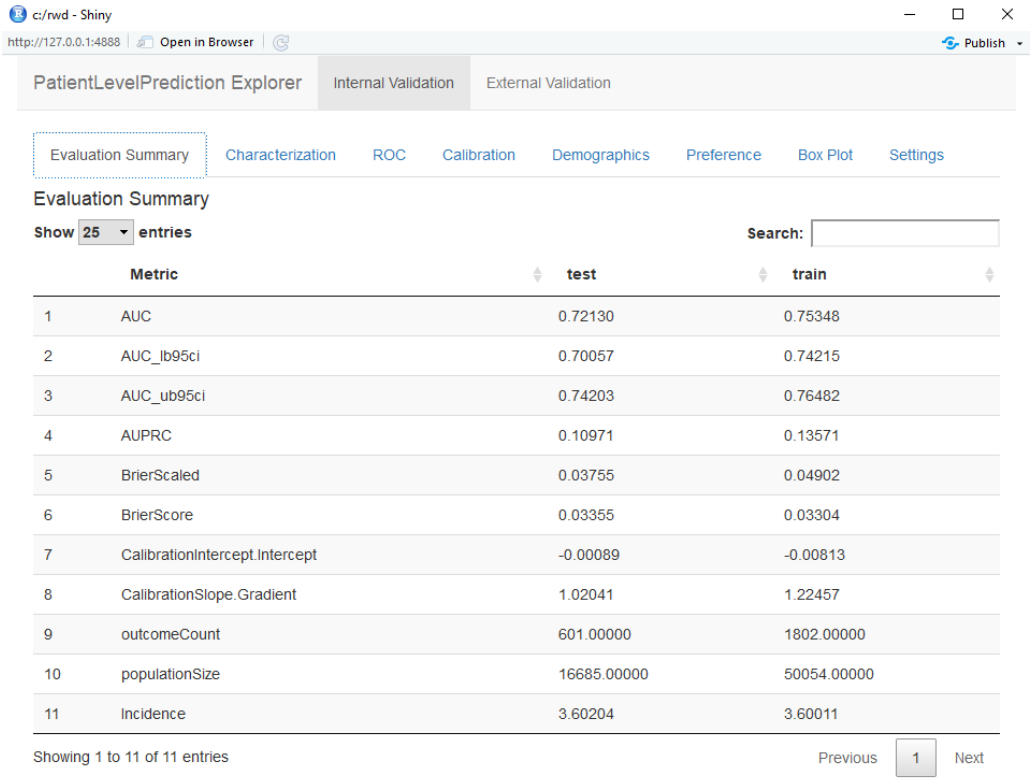
```
plpResult <- loadPlpResult(file.path(outputFolder,
                                     'Analysis_1',
                                     'plpResult'))
```

□□□□Analysis_1□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□Shiny□□□□□□□□□□□□□□□□□□□

```
viewPlp(plpResult)
```

Shiny□□□□□□□□□□□□□□□□□□□



	Metric	test	train
1	AUC	0.72130	0.75348
2	AUC_lb95ci	0.70057	0.74215
3	AUC_ub95ci	0.74203	0.76482
4	AUPRC	0.10971	0.13571
5	BrierScaled	0.03755	0.04902
6	BrierScore	0.03355	0.03304
7	CalibrationIntercept.Intercept	-0.00089	-0.00813
8	CalibrationSlope.Gradient	1.02041	1.22457
9	outcomeCount	601.00000	1802.00000
10	populationSize	16685.00000	50054.00000
11	Incidence	3.60204	3.60011

Figure 13.16: Shiny□□□□□□□

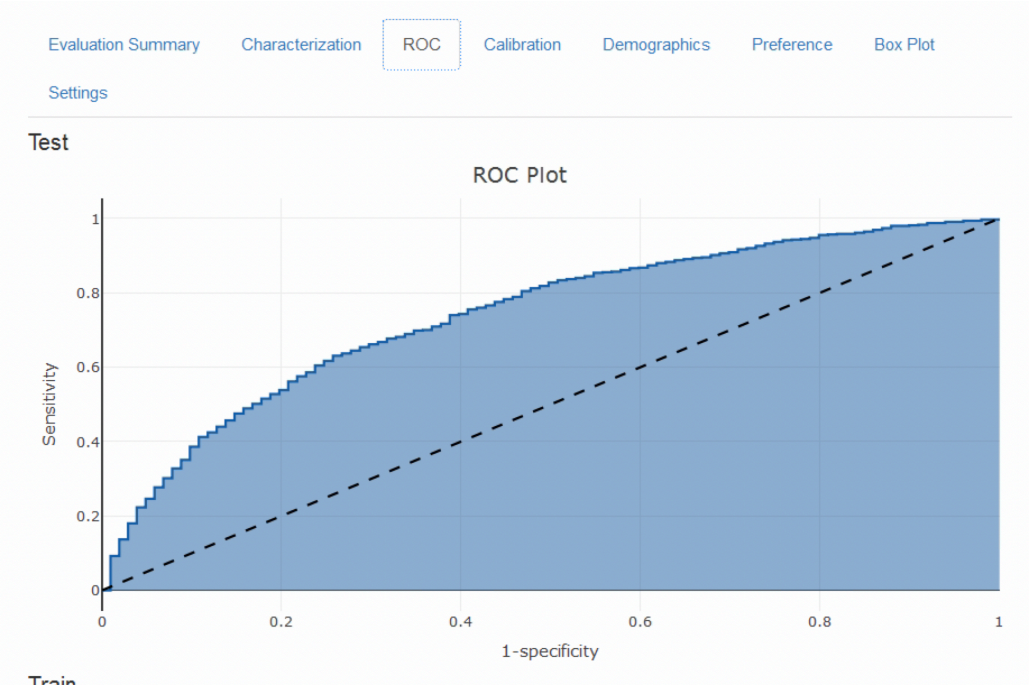


Figure 13.17: ROC□□□□□

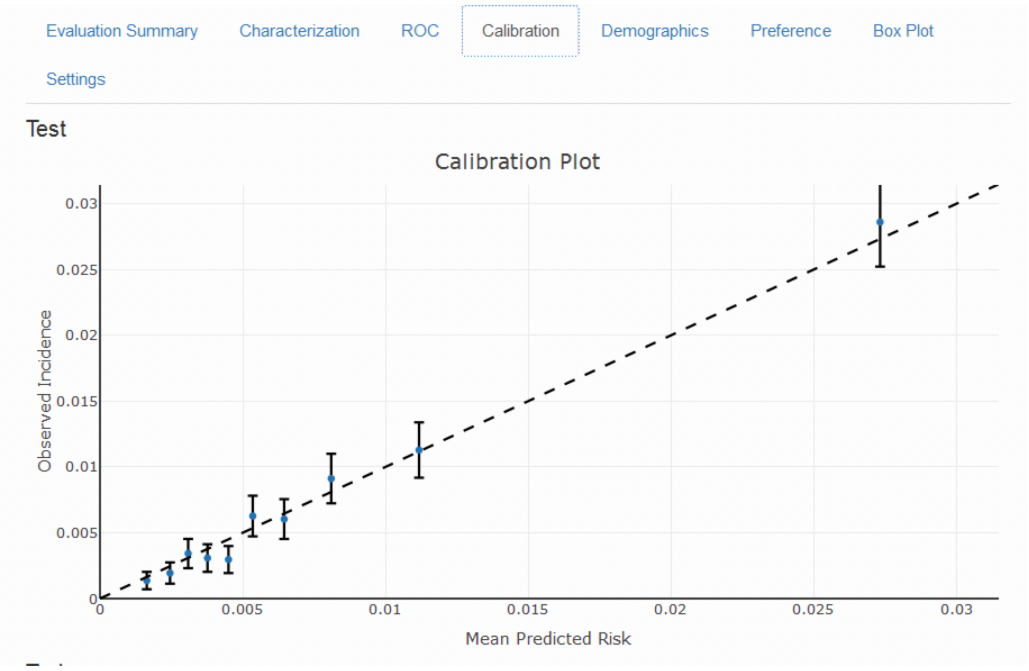


Figure 13.18: □□□□□□□□□□□□□□

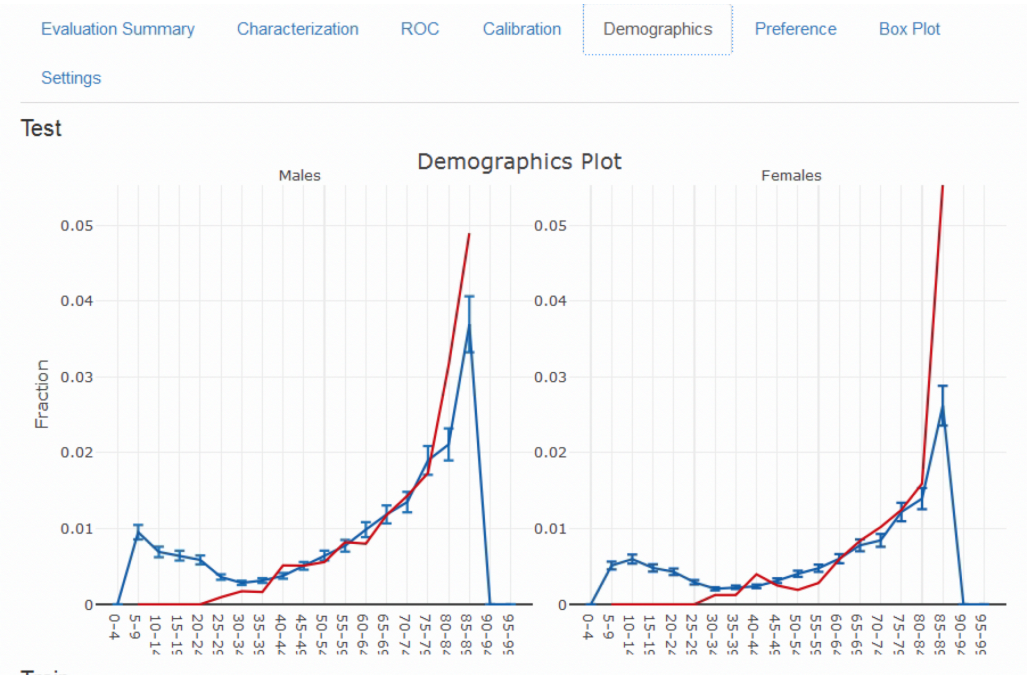


Figure 13.19: □□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□attrition□□□□□□□□□□□□□
13.20□□□□□□□□□□□□□□□□□□□

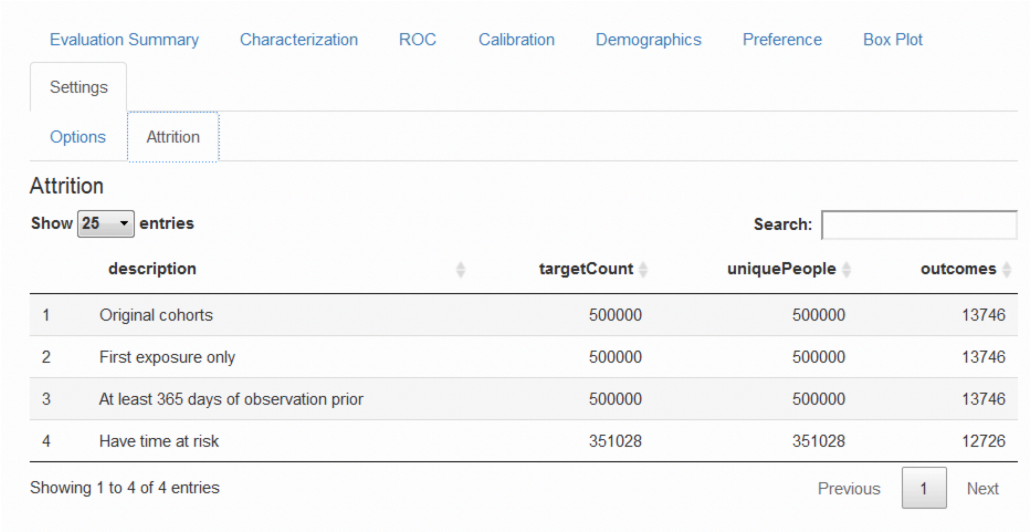


Figure 13.20: □□□□□□□□□attrition□□□□

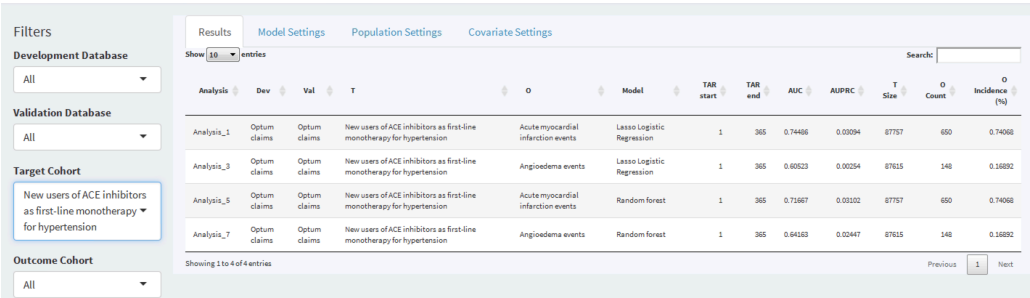


Figure 13.21: □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Shiny□□□□□□

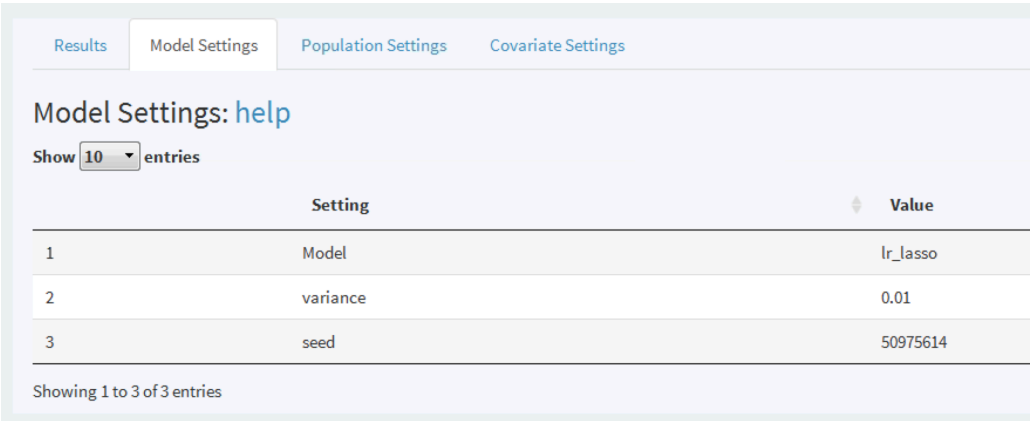


Figure 13.22: □□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□

□□□□□□□□□□□□□

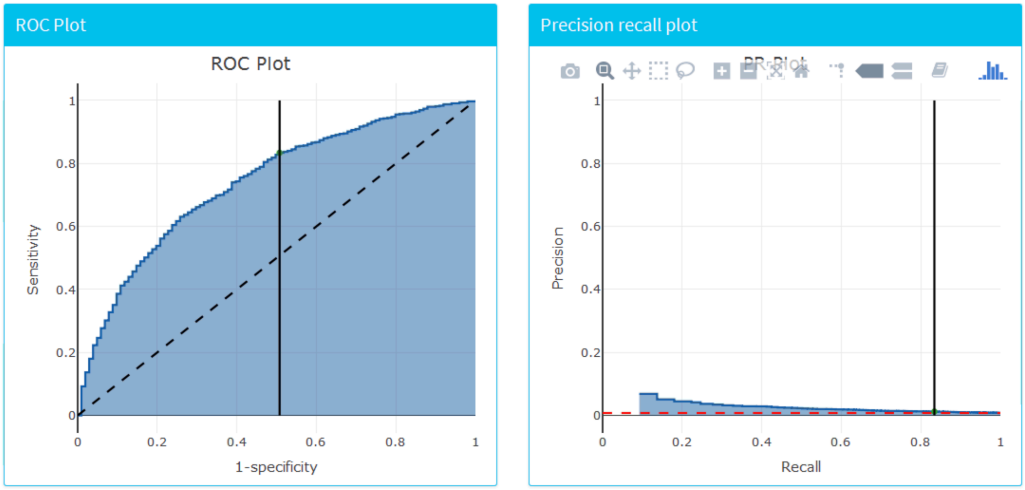


Figure 13.24: ROC□□□□□□-□□□□□□

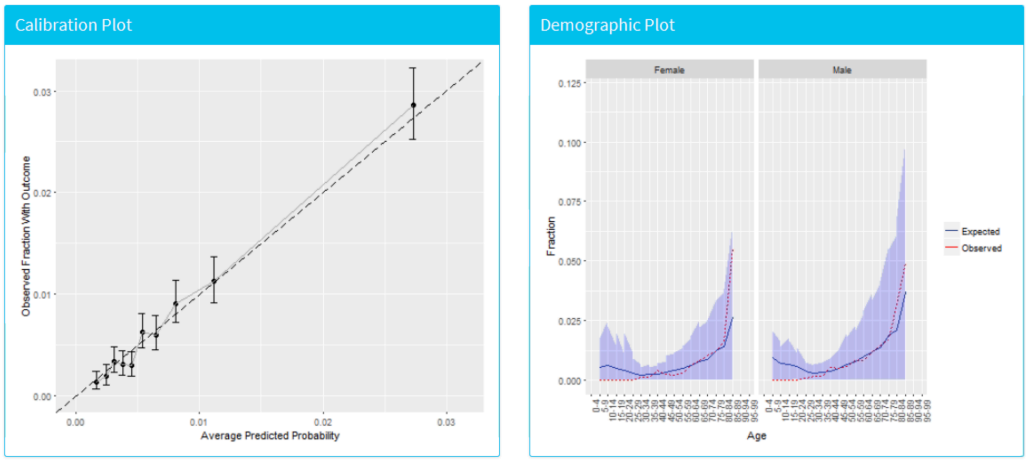

Calibration Plot
Observed Fraction With Outcome
Average Predicted Probability
Demographic Plot
Female
Male
Fraction
Age
Expected
Observed

Figure 13.26: □□□□□□□□

Model Table

Download Model

Show 10 entries

Search:

	Covariate Name	Value	Outcome Mean	Non-outcome Mean
1	age group: 00-04	0	0.0004	0.0001
2	age group: 05-09	0	0	0.0003
3	index month: 1	0	0.1307	0.1096
4	observation during day -365 through 0 days relative to index: Domain	0	0.1188	0.0514
5	Charlson index - Romano adaptation	0	2.4783	1.3817
6	Diabetes Comorbidity Severity Index (DCSI)	0.1478	2.4056	1.2207
7	CHADS2VASc	0.9279	3.0573	2.2576
8	visit_occurrence concept count during day -365 through 0 concept_count relative to index	0	19.5263	13.8837
9	age group: 10-14	0	0	0.001
10	index month: 2	0	0.0934	0.0909

Showing 1 to 10 of 67,897 entries Previous 1 2 3 4 5 ... 6790 Next

Figure 13.28: □□□□□□□□□□□

□□□□□□□□□□□□□

```
install.packages(c("SqlRender", "DatabaseConnector", "remotes"))
remotes::install_github("ohdsi/Eunomia", ref = "v1.0.0")
remotes::install_github("ohdsi/PatientLevelPrediction")
```

Eunomia□□□□□□□□CDM□□□□□□□□

(PART)

Chapter 14

□□□□: *Patrick Ryan & Jon Duke*

14.1

□□□□

14.2

□□□□□□□

Chapter 15

— —

□□□□*:* □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□

□□□□3□□□□□□□□□□□□□□□□□□□□□□□□□□□OHDSI□□□□□□□□□□□□□□□□□□□

Table 15.1: □□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□	□□□□□□□□	□□	□□
0.34	VISIT_OCCURRENCE □ provider_id □□□□□□□□□□□□□□□□□□□□□□□□□□□□□ yes□ no □□□	0.05	FAIL
0.99	MEASUREMENT □□□□□ measurement_source_value □□□□□□□□□□□□□□□□□□□□□□□□□□□ 0 □□□□□□□□□□□□□	0.30	FAIL
0.09	DRUG_ERA □□□□□ drug_concept_id □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□	0.10	PASS
0.02	DRUG_EXPOSURE □□□□□ DRUG_EXPOSURE_END_DATE □□□□□□□□□ DRUG_EXPOSURE_START_DATE □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□	0.05	PASS
0.00	PROCEDURE_OCCURRENCE □□□□□ procedure_occurrence_id □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□	0.00	PASS

□□□□□□□□□□□□□□□□□□□

```
source("Framework.R")
declareTest(101, "Person gender mappings")
add_enrollment(member_id = "M000000102", gender_of_member = "male")
add_enrollment(member_id = "M000000103", gender_of_member = "female")
expect_person(PERSON_ID = 102, GENDER_CONCEPT_ID = 8507
expect_person(PERSON_ID = 103, GENDER_CONCEPT_ID = 8532)
```

□□□□□□Rabbit-in-a-Hat□□□□□□

ID	□□	□□
101	Person gender mappings	PASS

□□□□□□□□□□□□□□ETL □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

15.3

□□□□□□□□□□□□□□□□□DQ□□□□□□□□□□□□□□□□□□□

□□□□□

□□□□□□resultsDatabaseSchema□□□□□□□□□□□□□□□□□□□□□□□□□CDM□□□□□□□□□□
ACHILLES□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□ATLAS□ACHILLES□□□□□□□□□

```
exportToJson(connectionDetails,
             cdmDatabaseSchema = cdmDatabaseSchema,
             resultsDatabaseSchema = cdmDatabaseSchema,
             outputPath = "achillesOut")
```

JSON□□□□□achillesOut□□□□□□□□□□□□□□□AchillesWeb□□□□□□□□□□□□□□□□□□□□□□

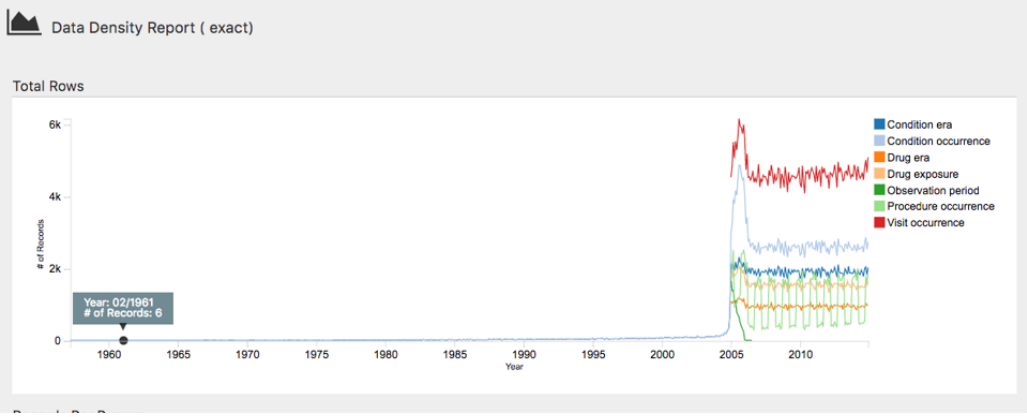


Figure 15.4: The data density plot in the ACHILLES web viewer.

□□□□□15.5□□□□□□□□□□□□□□□□□□□

□□□□□□DQD□CDM v5.3.1□□□□□□

RESULTS

SYNTHEA SYNTHETIC HEALTH DATABASE

Results generated at 2019-08-22 14:15:06 in 29 mins

Column visibility CSV

Show 5 entries Search:

STATUS	CONTEXT	CATEGORY	SUBCATEGORY	LEVEL	DESCRIPTION	% RECORDS
FAIL	Verification	Plausibility	Atemporal	FIELD	The number and percent of records with a value in the gap_days field of the DRUG_ERA table less than 0. (Threshold=0%).	24.07%
FAIL	Verification	Completeness	None	FIELD	The number and percent of records with a value of 0 in the standard concept field race_concept_id in the PERSON table. (Threshold=0%).	16.74%
FAIL	Verification	Conformance	Relational	FIELD	The number and percent of records that have a value in the ethnicity_concept_id field in the PERSON table that does not exist in the CONCEPT table. (Threshold=0%).	16.15%
PASS	Verification	Completeness	None	FIELD	The number and percent of records with a NULL value in the condition_end_date of the CONDITION_OCCURRENCE. (Threshold=100%).	13.24%
PASS	Verification	Completeness	None	FIELD	The number and percent of records with a NULL value in the condition_end_datetime of the CONDITION_OCCURRENCE. (Threshold=100%).	13.24%

Showing 71 to 75 of 1,327 entries (filtered from 1,639 total entries) Previous 1 ... 14 15 16 ... 266 Next

Figure 15.7: Data Quality Dashboard□□□□□□□□□□□□□□□□□□

15.6

□□□□□B.4□□□□□□□□□□□□□□□□□□

‰ per month	Max monthly ‰	Person count	Description
	60.60	24,189,656	**Inpatient or ER visit**
	39.50	15,003,249	**Emergency Room Visit** 9203
	39.50	15,003,249	ER (None) No matching concept
	23.90	9,186,407	**Inpatient Visit** 9201
	23.90	9,186,407	IP (None) No matching concept
	0.27	76,711	**Angioedema**
	0.27	76,711	**Angioedema** 432791
	0.26	64,726	9951 (ICD9CM) Angioneurotic edema, not elsewhere classified
	0.20	8,822	T783XXA (ICD10CM) Angioneurotic edema, initial encounter
	0.09	3,163	T783XXD (ICD10CM) Angioneurotic edema, subsequent encounter

Figure 15.8: □□□□□□□□□□□□□□□□□□□□□□□□□□□□

```
                                                    "Giant urticaria",
                                                    "Periodic edema"))
View(orphans)
```

□□□	□□	□□ID	□□□□□□□□
T78.3XXS	Angioneurotic edema, sequela	ICD10CM	508
10002425	Angioedemas	MedDRA	0
148774	Angioneurotic Edema of Larynx	CIEL	0
402383003	Idiopathic urticaria and/or angioedema	SNOMED	0
232437009	Angioneurotic edema of larynx	SNOMED	0
10002472	Angioneurotic edema, not elsewhere classified	MedDRA	0

□□□□□□□□□□□□□□□□□□□

15.7

- □□□□

Chapter 16

— —

□□□□: *Joel Swerdel, Seng Chan You, Ray Chen & Patrick Ryan*

□□□□

		Gold Standard	
		True	False
Cohort Definition	True	True Positive	False Positive
	False	False Negative	True Negative

Figure 16.1: □□□□□□□□□□□□□□□

□□□□□□□□□□□□□□

□□□PPV□□□□□□□□□□□□□□□□□□□

□□□□□□

□□□□ 2: xSens □□□□□□□□

□□□□□□□□□□□□□□□□□□xSens□□□□□□□□□□□□□□□□□□MI□□□□□□□□□□□□□□□□

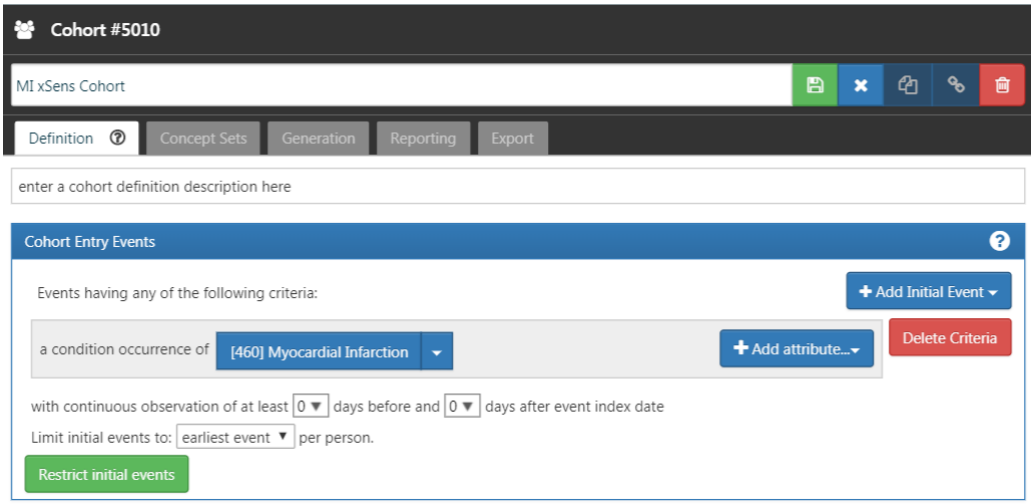


Figure 16.3: □□□□□□□□□□□□□□□□□□□□□□□xSens□

□□□□ 3: □□□□□□□□□

□□createPhenoModel□□□□□□□□□□□□□□□□□□□

```
  trainOutFile = "5XMI_train",
  exclCohort = 1770120, #xSens
  prevCohort = 1770119, #
  modelAnalysisId = "20181206V1",
  excludedConcepts = c(312327, 314666),
  addDescendantsToExclude = TRUE,
  cdmShortName = "myCDM",
  mainPopnCohort = 0, #
  lowerAgeLimit = 18,
  upperAgeLimit = 90,
  gender = c(8507, 8532),
  startDate = "20100101",
  endDate = "20171231")
```

□□□□□□□my_results□□□□□□□

```
  upperAgeLimit = 90,
  gender = c(8507, 8532),
  startDate = "20100101",
  endDate = "20171231")
```

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□c:/temp/lr_results_5XMI_train_myCDM_ePPV0
□□□□□□□□□□□□□□□□□□□

A

PersonId
016
019
022
023
025

Cohort according to definition to evaluate

PersonId	p(O)
016	0.99
017	0.01
018	0.00
019	0.72
020	0.02
021	0.00
022	0.15
023	0.22
024	0.00
025	0.00

Predicted probabilities

B

PersonId	p(O)	True Positive	False Positive	True Negative	False Negative
016	0.99	0.99	0.01		
017	0.01			0.99	0.01
018	0.00			1.00	0.00
019	0.72	0.72	0.28		
020	0.02			0.98	0.02
021	0.00			1.00	0.00
022	0.15	0.15	0.85		
023	0.22	0.22	0.78		
024	0.00			1.00	0.00
025	0.00	0.00	1.00		
	Sum	2.08	2.92	4.97	0.03

Computation of (weighted) confusion matrix

Sensitivity =	TP/(TP + FN) =	2.08/(2.08 + 0.03) =	0.99
Specificity =	TN/(TN + FP) =	4.97/(4.97 + 2.92) =	0.63
Positive Predictive Value =	TP/(TP + FP) =	2.08/(2.08 + 2.92) =	0.42
Negative Predictive Value =	TN/(TN + FN) =	4.97/(4.97 + 0.03) =	0.99

C Computation of cohort evaluation metrics

Figure 16.4: Determining the Performance Characteristics of a cohort definition using PheValuator. p(O) = Probability of outcome; TP = True Positive; FN = False Negative; TN = True Negative; FP = False Positive.

```
  password = "supersecret")

phenoResult <- testPhenotype(
  connectionDetails = connectionDetails,
  cutPoints = c(0.1, 0.2, 0.3, 0.4, 0.5, "EV", 0.6, 0.7, 0.8, 0.9),
  resultsFileName =
    "c:/temp/lr_results_5XMI_eval_myCDM_ePPV0.75_20181206V1.rds",
  modelFileName =
    "c:/temp/lr_results_5XMI_train_myCDM_ePPV0.75_20181206V1.rds",
  cohortPheno = 1769702,
  phenText = "All MI by Phenotype 1 X In-patient, 1st Position",
  order = 1,
  testText = "MI xSpec Model - 5 X MI",
  cohortDatabaseSchema = "my_results",
  cohortTable = "cohort",
  cdmShortName = "myCDM")
```

□□□□□□□□□□□□□□□□□□□cutPoints□□□□□□□□□□□□□□□□□EV□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□16.1□□5□□□□□□□□□□□□□MI□4□□□□□□□□□□□□□□□□□□□

Phenotype Algorithm	Database	Sens	PPV	Spec	NPV

16.5

□□□□□□

Chapter 17

— —

□□□□: *Martijn Schuemie*

□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□

- □□□□□□□□□□□

- □□□□□□□□□□□□□□□□4.3.3 □□ 4.4.0□□□□□□□□□□□□□□□

17.3.2

□□□□□□□□□□

Chapter 18

— —

□□□□: *Martijn Schuemie*

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□

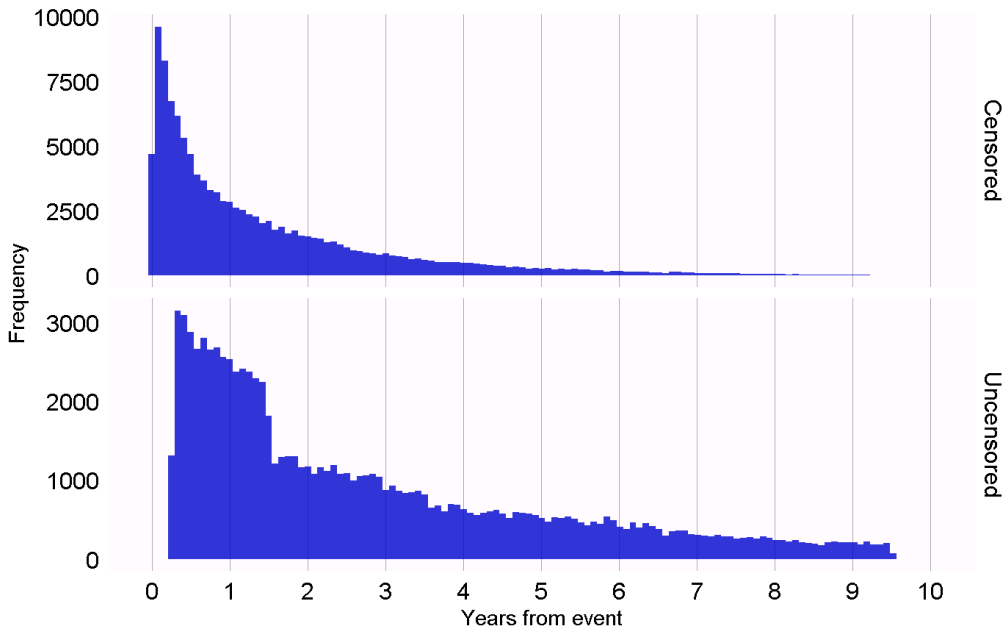

10000
7500
5000
2500
0
Frequency
3000
2000
1000
0
Censored
Uncensored
0
1
2
3
4
5
6
7
8
9
10
Years from event

Figure 18.1: □□□□□□

Tchetgen Tchetgen, and Cohen 2010□□□□□□□□□

18.2.4 P

□□□□□□1□□□□□α=0.05□□5%□□□□□□

□□□□□

$$\mu(\theta_i) = a + b \times \theta_i \text{ and}$$
$$\sigma(\theta_i)^2 = c + d \times \mid \theta_i \mid$$

a b c d beta_i$□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

$$l(a,b,c,d|\theta,\hat{\theta},\hat{\tau}) \propto \prod_{i=1}^{n} \int p(\hat{\theta}_i|\beta_i,\theta_i,\hat{\tau}_i)p(\beta_i|a,b,c,d,\theta_i)\mathrm{d}\beta_i,$$

□□□□□ $(\hat{a},\ b,\ c,\ d)$.

□□□□□□□□□□□□□□□□□□□

18.3

□□□□□□□12□□□□□□□□□ACE□□□□□ACEi□□□□□□□□□□□□□□□□□□□□□AMI□□□□□□□□□□□□□

18.3.1

□□□□□□□□□□□□□□□□□□□

Evidence for all conditions for ACEi and THZ combined

Save New Concept Set From Selection Below | View database record counts (RC) and descendant record counts (DRC) for: SYNPUF 5%

Column visibility | Copy | CSV | Show 15 entries | Filter:

Showing 1 to 15 of 13,787 entries | Previous 1 2 3 4 5 ... 920 Next

Suggested Negative Control
No (12777)
Yes (1010)

Found in Publications
No (12398)
Yes (Parent) (1160)
Yes (Exact) (229)

Found on Product Label
No (12667)
Yes (Parent) (878)
Yes (Exact) (242)

Found in Product Label Or Publications
Yes (10576)
No (3211)

Signal in FAERS
No (10951)
Yes (Parent) (1949)

Name	Suggested Negative Control	Sort Order	Publication Count (Descendant Concept Match)	Publication Count (Exact Concept Match)	Publication Count (Parent Concept Match)	Product Label Count (Descendant Concept Match)	Product Label (Exact Concept Match)	Product Label (Parent Concept Match)
Rift valley fever	Y	13,781	0	0	0	0	0	0
Obstruction due to foreign body accidentally left in operative wound AND/OR body cavity during a procedure	Y	13,780	0	0	0	0	0	0
Infection by Shigella	Y	13,766	0	0	0	0	0	0

Figure 18.4: □□□□□□□□□□□□□□

□□□□□□

```
                    seLogRrNegatives = ncEstimates$seLogRr,
                    logRrPositives = oiEstimates$logRr,
                    seLogRrPositives = oiEstimates$seLogRr,
                    showCis = TRUE)
```

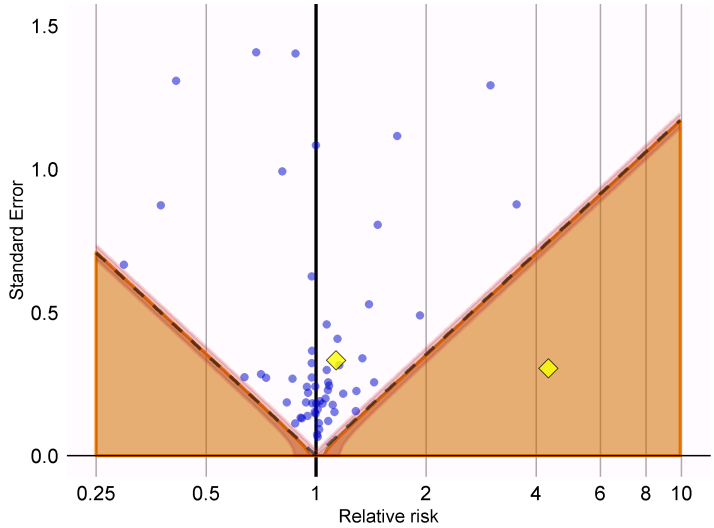


Figure 18.6: P-value calibration: estimates below the dashed line have a conventional p < 0.05. Estimates in the shaded area have calibrated p < 0.05. The narrow band around the edge of the shaded area denotes the 95% credible interval. Dots indicate negative controls. Diamonds indicate outcomes of interest.

□18.6□□□□□□□□□□□□□

18.3.5

□□□□□

18.3.7

□□□□□□□□□□□□□□

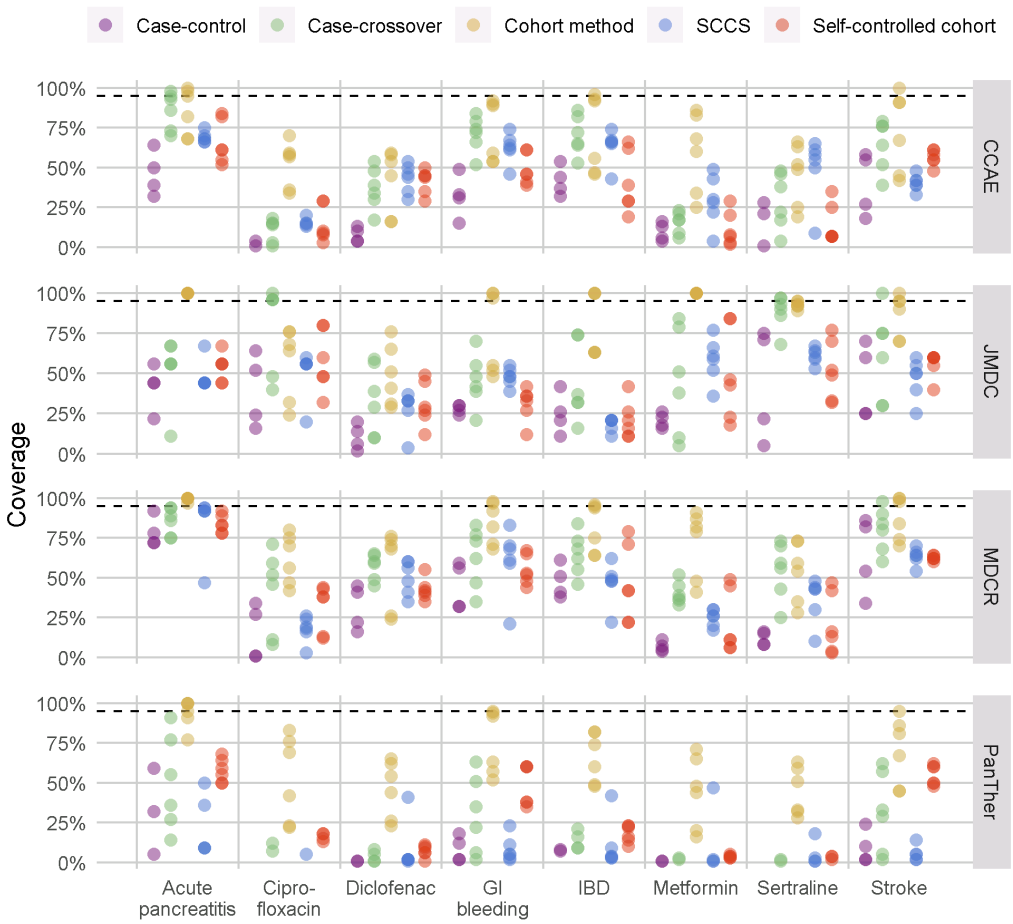


Figure 18.9: Methods Library□Methods□□□□95%□□□□□

Part IV

OHDSI

Chapter 19

— —

□□□□: *Sara Dempster & Martijn Schuemie*

□□□□□OHDSI□□□□□

□□□□□

19.2.3

□□□□□□□

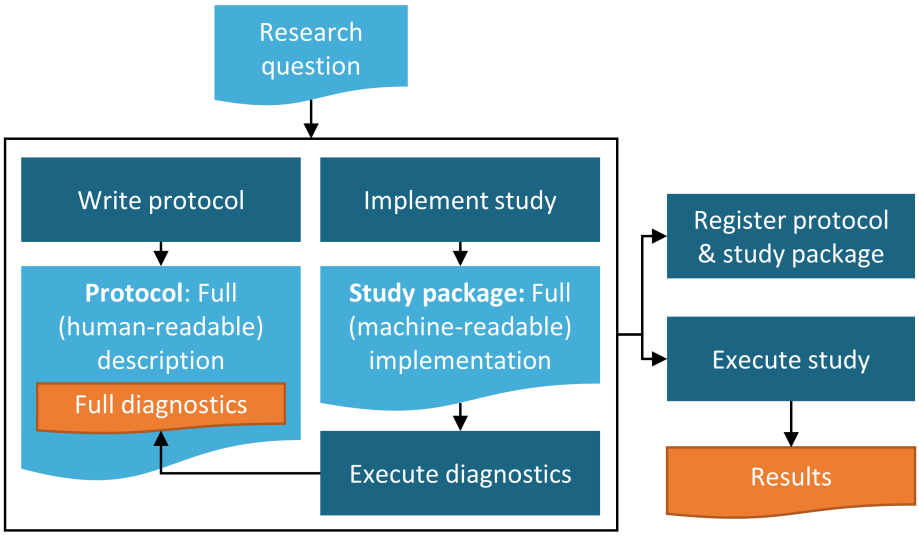


Figure 19.1: □□□□□□□□□□□□□□□

19.2.6

□□□□□□□□□□□□□□

Chapter 20

– – OHDSI

□□□□: *Kristin Kostka, Greg Klebanov & Sara Dempster*

OHDSI□□□□

20.2.2 OHDSI

□□□□□□□□

- □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
- □□□□□CDM□□
- □□□□□□□□□□□□□□□□□□□□□□RStudio□□□□□□□□□□□□
- □□□□□□□□□□□□□□□□□□□□□□□
- □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□R□□□□□□□□□□□□□□□□□□□

**□□□□□□□□□□□□□□□□□□□6□□□□□□□□□□□□□□□□□ETL□□□□□□□□□□□□□□□□□□□

□□□□□□□□

ARACHNE□□□□□□□□□□□□□□□□□□□

- □□□□□□□□□□□□□□□□□□□□□□OHDSI□□□□□□□□□□□□□□□□□□□

Chapter A

— —

ACHILLES □□□□□□□□□□□□□□□□□□□□□□

ARACHNE □□OH

ATLAS □□

□□□□□**Bias**□ □□□□□□□□□□□□□□□□□□□□

□□□□□□□□**Boolean**□ 2□□□□□□□□□□□□□□□□□□□□

□□□□□**Care site**□ □□□

□□□□□□□□□**Case control**□ □□□

□□□□□**Causal effect**□ □□

□□□□□□**Characterization**□ □□□□□□□□□□□□□□□□□□□□

□□□□□□**Data quality**□ □□□□□□□

Rabbit-in-a-Hat □□□□□□□□□□□□□□□□□□□□ETL□□□□□□□□□□□□□□□□□□□

Chapter B

□□□

Table B.1: ACE□□□

□□□□□ID	□□□□□□	□□	□□□	□□□□□□
1308216	□□□□□□	□□□	□□	□□□
1310756	□□□□□□□	□□□	□□	□□□
1331235	□□□□□	□□□	□□	□□□
1334456	□□□□□	□□□	□□	□□□
1335471	□□□□□□	□□□	□□	□□□
1340128	□□□□□□	□□□	□□	□□□
1341927	□□□□□□	□□□	□□	□□□
1342439	□□□□□□□□	□□□	□□	□□□
1363749	□□□□□□□	□□□	□□	□□□
1373225	□□□□□□□	□□□	□□	□□□

B.2 ACE

□□□□□□□□□□□

□□□□□□□□□□□□□

- □□□□□□□□□□□□□□□*ACE*□□□□□□ B.2□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□365□□□□□□□□□□□□□0□□□□□□□□□□□□□□□□□□□

□□□□□□□□

□□□□□□□□□□□□□□

□□□□□ID	□□□□□□	□□	□□□	□□□□□□
974166	□□□□□□□□□□	□□□	□□	□□□
978555	□□□□□□	□□□	□□	□□□
991382	□□□□□	□□□	□□	□□□
1305447	□□□□□	□□□	□□	□□□
1307046	□□□□□□□	□□□	□□	□□□
1307863	□□□□□	□□□	□□	□□□
1308216	□□□□□□	□□□	□□	□□□
1308842	□□□□□□	□□□	□□	□□□
1309068	□□□□□□	□□□	□□	□□□
1309799	□□□□□□	□□□	□□	□□□
1310756	□□□□□□□	□□□	□□	□□□
1313200	□□□□□	□□□	□□	□□□
1314002	□□□□□□	□□□	□□	□□□
1314577	□□□□□□	□□□	□□	□□□
1317640	□□□□□□□	□□□	□□	□□□
1317967	□□□□□□	□□□	□□	□□□
1318137	□□□□□□	□□□	□□	□□□
1318853	□□□□□□	□□□	□□	□□□
1319880	□□□□□□	□□□	□□	□□□
1319998	□□□□□□□	□□□	□□	□□□
1322081	□□□□□□□	□□□	□□	□□□
1326012	□□□□□□	□□□	□□	□□□
1327978	□□□□□□□	□□□	□□	□□□
1328165	□□□□□□	□□□	□□	□□□
1331235	□□□□□	□□□	□□	□□□
1332418	□□□□□□	□□□	□□	□□□
1334456	□□□□□	□□□	□□	□□□
1335471	□□□□□□	□□□	□□	□□□
1338005	□□□□□□□	□□□	□□	□□□
1340128	□□□□□□	□□□	□□	□□□
1341238	□□□□□	□□□	□□	□□□
1341927	□□□□□□	□□□	□□	□□□
1342439	□□□□□□□□	□□□	□□	□□□
1344965	□□□□□□□	□□□	□□	□□□
1345858	□□□□□□	□□□	□□	□□□
1346686	□□□□□□□	□□□	□□	□□□
1346823	□□□□□□□	□□□	□□	□□□
1347384	□□□□□□□	□□□	□□	□□□
1350489	□□□□□	□□□	□□	□□□
1351557	□□□□□□□	□□□	□□	□□□
1353766	□□□□□□□□	□□□	□□	□□□
1353776	□□□□□□	□□□	□□	□□□
1363053	□□□□□□	□□□	□□	□□□

□□□□□ID	□□□□□□□	□□	□□□	□□□□□□□
1363749	□□□□□□□□	□□□	□□	□□□
1367500	□□□□□□	□□□	□□	□□□
1373225	□□□□□□□□	□□□	□□	□□□
1373928	□□□□□□□	□□□	□□	□□□
1386957	□□□□□□□	□□□	□□	□□□
1395058	□□□□□□□□	□□□	□□	□□□
1398937	□□□□□□	□□□	□□	□□□
40226742	□□□□□□□□	□□□	□□	□□□
40235485	□□□□□□□□	□□□	□□	□□□

B.3 AMI

□□□□□□□□□□□

□□□□□□□□□□□□□□

- □□□□□□□□□ B.5□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□0□□□□□□□□□□□□□0□□□□□□□□□□□□□□□□□□□

Table B.6: □□□□□□□□□□□

□□□□□ID	□□□□□□□	□□	□□□	□□□□□□□
262	□□□□□□□□□□□□□	□□□	□□	□□□
9201	□□□□□□□	□□□	□□	□□□
9203	□□□□□□□□	□□□	□□	□□□

B.4

□□□□□□□□□□□

□□□□□□□□□□□□□

- □□□□□□□□□□□□□□□□□□□□ B.7□

□□□□□□□□□□□□□□□□□□□0□□□□□□□□□□□□□□□□□□□

□□□□□□ID	□□□□□□□	□□	□□□	□□□□□□□
9201	□□□□□□□	□□□	□□	□□□
9203	□□□□□□□□	□□□	□□	□□□

B.5

□□□□□□□□□□□

□□□□□□□□□□□□□

- □□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□

Table B.9: □□□□□□□□□□□□□□□□□□□□□

□□□□□ID	□□□□□□	□□	□□□	□□□□□□
907013	□□□□□	□□□	□□	□□□
974166	□□□□□□□□□□	□□□	□□	□□□
978555	□□□□□□	□□□	□□	□□□
1395058	□□□□□□□	□□□	□□	□□□

Table B.10: □□□□□□

□□□□□ID	□□□□□□□	□□	□□□	□□□□□□□
316866	□□□□□□□	□□□	□□	□□□

Table B.11: □□□□□□

□□□□□ID	□□□□□□	□□	□□□	□□□□□□
904542	□□□□□□□	□□□	□□	□□□
907013	□□□□□	□□□	□□	□□□
932745	□□□□□	□□□	□□	□□□
942350	□□□□□	□□□	□□	□□□
956874	□□□□□	□□□	□□	□□□
970250	□□□□□□□□	□□□	□□	□□□
974166	□□□□□□□□□□	□□□	□□	□□□
978555	□□□□□□	□□□	□□	□□□
991382	□□□□□□	□□□	□□	□□□
1305447	□□□□□	□□□	□□	□□□
1307046	□□□□□□□	□□□	□□	□□□
1307863	□□□□□	□□□	□□	□□□
1308216	□□□□□□	□□□	□□	□□□
1308842	□□□□□□	□□□	□□	□□□
1309068	□□□□□□	□□□	□□	□□□
1309799	□□□□□□	□□□	□□	□□□
1310756	□□□□□□□	□□□	□□	□□□
1313200	□□□□□	□□□	□□	□□□
1314002	□□□□□□	□□□	□□	□□□
1314577	□□□□□□	□□□	□□	□□□
1317640	□□□□□□□	□□□	□□	□□□
1317967	□□□□□□	□□□	□□	□□□
1318137	□□□□□□□	□□□	□□	□□□
1318853	□□□□□□	□□□	□□	□□□

□□□□□ID	□□□□□□□	□□	□□□	□□□□□□□
1319880	□□□□□□□	□□□	□□	□□□
1319998	□□□□□□□□	□□□	□□	□□□
1322081	□□□□□□□	□□□	□□	□□□
1326012	□□□□□□□	□□□	□□	□□□
1327978	□□□□□□□□	□□□	□□	□□□
1328165	□□□□□□□	□□□	□□	□□□
1331235	□□□□□□	□□□	□□	□□□
1332418	□□□□□□□	□□□	□□	□□□
1334456	□□□□□□	□□□	□□	□□□
1335471	□□□□□□□	□□□	□□	□□□
1338005	□□□□□□□□	□□□	□□	□□□
1340128	□□□□□□□	□□□	□□	□□□
1341238	□□□□□□	□□□	□□	□□□
1341927	□□□□□□□	□□□	□□	□□□
1342439	□□□□□□□□□	□□□	□□	□□□
1344965	□□□□□□□□	□□□	□□	□□□
1345858	□□□□□□□	□□□	□□	□□□
1346686	□□□□□□□□	□□□	□□	□□□
1346823	□□□□□□□□	□□□	□□	□□□
1347384	□□□□□□□□	□□□	□□	□□□
1350489	□□□□□□	□□□	□□	□□□
1351557	□□□□□□□□	□□□	□□	□□□
1353766	□□□□□□□□□	□□□	□□	□□□
1353776	□□□□□□□	□□□	□□	□□□
1363053	□□□□□□□	□□□	□□	□□□
1363749	□□□□□□□□	□□□	□□	□□□
1367500	□□□□□□	□□□	□□	□□□
1373225	□□□□□□□□	□□□	□□	□□□
1373928	□□□□□□□	□□□	□□	□□□
1386957	□□□□□□□	□□□	□□	□□□
1395058	□□□□□□□□	□□□	□□	□□□
1398937	□□□□□□	□□□	□□	□□□
40226742	□□□□□□□□	□□□	□□	□□□
40235485	□□□□□□□□	□□□	□□	□□□

B.6

□□□□□□□□□□□

□□□□□□□□□□□□□

- □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□ B.12□□□□□□

□□□□□□□□□□□□□□□□□□□□□□365□□□□□□□□□□□□365□□□□□□□□□□□□□□□□□□□□□□

□□□□□

□□□□□□□□□□□□□□□□

- □□□□□□□□□□1□□□□□□□□□□□□□□□□□1□□□□□□□□□□□ B.13□□□□□□□□□□□□□□
- □□□□□□□□□□□□□□□□□□□□□□365□□□□0□□□□□□□□□□□□□□□□□
  B.14□□□□□□□□□□□□□□□□□□□□1□□□□□□□□

□□□□□□□□□□□□□□□□□□□

□□□□□ID	□□□□□□□	□□	□□□	□□□□□□□
1347384	□□□□□□□□	NO	YES	NO
1351557	□□□□□□□□	NO	YES	NO
1353776	□□□□□□□	NO	YES	NO
1363749	□□□□□□□	NO	YES	NO
1367500	□□□□□□	NO	YES	NO
1373225	□□□□□□□□	NO	YES	NO
1395058	□□□□□□□□	NO	YES	NO
40226742	□□□□□□□□	NO	YES	NO
40235485	□□□□□□□□	NO	YES	NO

Table B.13: □□□□□□

□□□□□ID	□□□□□□□	□□	□□□	□□□□□□□
904542	□□□□□□□□	NO	YES	NO
907013	□□□□□□	NO	YES	NO
932745	□□□□□□	NO	YES	NO
942350	□□□□□□	NO	YES	NO
956874	□□□□□□	NO	YES	NO
970250	□□□□□□□□□	NO	YES	NO
974166	□□□□□□□□□□□	NO	YES	NO
978555	□□□□□□□	NO	YES	NO
991382	□□□□□□	NO	YES	NO
1305447	□□□□□□	NO	YES	NO
1307046	□□□□□□□□	NO	YES	NO
1307863	□□□□□□	NO	YES	NO
1308216	□□□□□□□	NO	YES	NO
1308842	□□□□□□□	NO	YES	NO
1309068	□□□□□□□	NO	YES	NO
1309799	□□□□□□□	NO	YES	NO
1310756	□□□□□□□□	NO	YES	NO
1313200	□□□□□□	NO	YES	NO
1314002	□□□□□□□	NO	YES	NO
1314577	□□□□□□□	NO	YES	NO
1317640	□□□□□□□□	NO	YES	NO
1317967	□□□□□□□	NO	YES	NO
1318137	□□□□□□□	NO	YES	NO
1318853	□□□□□□□	NO	YES	NO
1319880	□□□□□□□	NO	YES	NO
1319998	□□□□□□□□	NO	YES	NO
1322081	□□□□□□□□	NO	YES	NO
1326012	□□□□□□□	NO	YES	NO
1327978	□□□□□□□□	NO	YES	NO

□□□□□ID	□□□□□□	□□	□□□	□□□□□□
1328165	□□□□□□	NO	YES	NO
1331235	□□□□□	NO	YES	NO
1332418	□□□□□□	NO	YES	NO
1334456	□□□□□	NO	YES	NO
1335471	□□□□□□	NO	YES	NO
1338005	□□□□□□□	NO	YES	NO
1340128	□□□□□□	NO	YES	NO
1341238	□□□□□	NO	YES	NO
1341927	□□□□□□	NO	YES	NO
1342439	□□□□□□□□	NO	YES	NO
1344965	□□□□□□□	NO	YES	NO
1345858	□□□□□□	NO	YES	NO
1346686	□□□□□□□	NO	YES	NO
1346823	□□□□□□□	NO	YES	NO
1347384	□□□□□□□	NO	YES	NO
1350489	□□□□□	NO	YES	NO
1351557	□□□□□□□	NO	YES	NO
1353766	□□□□□□□□	NO	YES	NO
1353776	□□□□□□	NO	YES	NO
1363053	□□□□□□	NO	YES	NO
1363749	□□□□□□	NO	YES	NO
1367500	□□□□□	NO	YES	NO
1373225	□□□□□□□	NO	YES	NO
1373928	□□□□□□	NO	YES	NO
1386957	□□□□□□	NO	YES	NO
1395058	□□□□□□□	NO	YES	NO
1398937	□□□□□	NO	YES	NO
40226742	□□□□□□□	NO	YES	NO
40235485	□□□□□□□	NO	YES	NO

Table B.14: □□□□□□

□□□□□ID	□□□□□□	□□	□□□	□□□□□□
316866	□□□□□□	NO	YES	NO

B.7 3

□□□□□□ B.6 □□□□□□□□□□□□□□□□□□□□□□□□□□□□365□□□□□**1095**□□□□□□□□□□□□□□□

B.8 ACE

□□□□□□□□□□□

□□□□□□□□□□□□□

- *ACE*□□□□□ B.15□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□0□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□

Table B.16: □□□□□□□□□□□□□□□□ARB□

□□□□□ID	□□□□□□	□□	□□□	□□□□□□
1308842	□□□□□□	NO	YES	NO
1317640	□□□□□□□	NO	YES	NO
1346686	□□□□□□□	NO	YES	NO
1347384	□□□□□□□	NO	YES	NO
1351557	□□□□□□□	NO	YES	NO
1367500	□□□□□	NO	YES	NO
40226742	□□□□□□□	NO	YES	NO
40235485	□□□□□□□	NO	YES	NO

B.10

□□□□□□ B.8 □□□□□□□□□□□□□□□□□□□□

□□□□□ID	□□□□□□	□□	□□□	□□□□□□
1332418	□□□□□□	NO	YES	NO
1353776	□□□□□□	NO	YES	NO

B.12 NDCCB

□□□□□□ B.8 □□□□□□□

B.14

ACE□□□□□ B.8 □□□□□□□□□□□□□□ (□ B.21) □*ACE*□□□ (□ B.15) □□□□□□□□□□□□□

□□□□□□□□□□□

Table B.21: □□□□□□

□□□□□ID	□□□□□□	□□	□□□	□□□□□□
932745	□□□□□	NO	YES	NO
942350	□□□□□	NO	YES	NO
956874	□□□□□	NO	YES	NO

B.15

ACE□□□□□ B.8 □□□□□□□□□□□□□□□□□□ (□ B.22) □ *ACE*□□□ (□ B.15) □□□□□□□□□□□□□

□□□□□□□□□□□

Table B.22: □□□□□□□□□□

□□□□□ID	□□□□□□	□□	□□□	□□□□□□
904542	□□□□□□□	NO	YES	NO
991382	□□□□□□	NO	YES	NO

B.16 1

ACE□□□□□ B.8 □□□□□□□□*α1*□□□ (□ B.23) □ *ACE*□□□ (□ B.15)□□□□□□□□□□□□□

□□□□□□□□□□□

Table B.23: α1□□□

□□□□□ID	□□□□□□	□□	□□□	□□□□□□
1341238	□□□□□	NO	YES	NO
1350489	□□□□□	NO	YES	NO
1363053	□□□□□□	NO	YES	NO

Chapter C

□□□

□□□□□ID	□□□□□□□□
134438	□□□□□□
78619	□□□□□
201606	□□□□□
76786	□□□□□□□
4115402	□□□□
45757370	□□□□□□□□
433111	□□□□□
433527	□□□□□
4170770	□□□□□
4092896	□□□□□□□
259995	□□□□□□□
40481632	□□□□□□□□
4166231	□□□□□
433577	□□□
4231770	□□□□□□
440329	□□□□□□□□□□□□
4012570	□□□□□□□□□
4012934	□□□□□□□□
441788	□□□□□□□□□□□□□□
4201717	□□□□□□
374375	□□□□
4344500	□□□□□□□□□□□□□□□
139099	□□□
444132	□□□
196168	□□□□
432593	□□□□□□□
434203	□□□□□□
438329	□□□□□□□□□
195873	□□□□
4083487	□□□□□□□
4103703	□□□□□□□□
4209423	□□□□□□□
377572	□□□□□□□□□□□
40480893	□□□□□□□□□□□□□□□
136368	□□□□□□□□□□□
140648	□□□□□□□□□□□
438130	□□□□□□□
4091513	□□
4202045	□□□□□□□□□□□□□
373478	□□
46286594	□□□□□□□□□□□□□□□
439790	□□□□
81634	□□□□

□□□□□ID	□□□□□□□□□
380706	□□□
141932	□□□□□□
36713918	□□□□□□□□□□
443172	□□□□□□□□□□□□□
81151	□□□□□
72748	□□□□□□□□□□□□
378427	□□□□
437264	□□□□□□□□□
194083	□□□□□□□□□□
140641	□□□□□
440193	□□□□
4115367	□□□□

Chapter D

1. □□
2. □□□□
3. □□
4. □□□□□
5. □□□□□□□
6. □□□□□□□□
7. □□□□
    - □□□□
    - □□□□
    - □□□□
    - □□□□
8. □□□□
    - □□□□□□
    - □□□□□□
    - □□□□□□
    - □□
    - □□□□□□□□□
    - □□□
9. □□□□□□□
    - □□□□□□□□
    - □□□□□
    - □□□□□□□□□□□□□□□
    - □□□□□□
    - □□
    - □□□□□□□
10. □□□□
    - □□□□□□□□□□□
    - □□□□□□□□□□
    - □□□□□□□□

11. □□□□□□□□□□□
12. □□□□□□□□□
13. □□□□□□□□□□□□□□□□□□□
14. □□□□□□□□□□□□□□□□□□□□□□□□
15. □□□□□□□□
16. □□□□

Chapter E

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

E.1

□□ 4.1

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□ E.1 □□□□□□□□□□□□

Table E.1: PERSON□□□□□

□□□□	□	□□
PERSON_ID	2	□□□□□□
GENDER_CONCEPT_ID	8507	□□□□□□□□ ID □ 8507□
YEAR_OF_BIRTH	1974	
MONTH_OF_BIRTH	8	
DAY_OF_BIRTH	4	
BIRTH_DATETIME	1974-08-04 00:00:00	□□□□□□□□□□0□□00:00:00□□□□
DEATH_DATETIME	NULL	
RACE_CONCEPT_ID	8516	□□□□□□□□□□□□□□□□□ ID □ 8516□
ETHNICITY_ CONCEPT_ID	38003564	38003564 □□□□□□□□□□□□□□□
LOCATION_ID		□□□□□□
PROVIDER_ID		□□□□□□□
CARE_SITE		□□□□□□□□□□□
PERSON_SOURCE_ VALUE	NULL	□□□□□□□□□□
GENDER_SOURCE_ VALUE	Man	□□□□□□□□□□□□□□

□□□□	□	□□
GENDER_SOURCE_CONCEPT_ID	0	
RACE_SOURCE_VALUE	African American	□□□□□□□□□□□□□□
RACE_SOURCE_CONCEPT_ID	0	
ETHNICITY_SOURCE_VALUE	NULL	
ETHNICITY_SOURCE_CONCEPT_ID	0	

□□ 4.2

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□ E.2 □□□□□□□□□□□□

Table E.2: OBSERVATION_PERIOD□□□□□

□□□□	□	□□
OBSERVATION_PERIOD_ID	2	□□□□□□
PERSON_ID	2	□□□PERSON□□□□□□□□□□□□□□□□□
OBSERVATION_PERIOD_START_DATE	2015-01-01	□□□□□□□□
OBSERVATION_PERIOD_END_DATE	2019-07-01	□□□□□□□□□□□□□□□□□□□□□□□□□□□□
PERIOD_TYPE_CONCEPT_ID	44814722	44814724 □□□□□□□□□□□□□□

□□ 4.3

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□ E.3 □□□□□□□□□□□□

Table E.3: DRUG_EXPOSURE□□□□□

□□□□	□	□□
DRUG_EXPOSURE_ID	1001	□□□□□□
PERSON_ID	2	□□□PERSON□□□□□□□□□□□□□□□□□
DRUG_CONCEPT_ID	19078461	□□□□□NDC□□□□□□□□□□□□ 19078461 □□□□□□□□□□□
DRUG_EXPOSURE_START_DATE	2019-05-01	□□□□□□□□□□□

□□□□	□	□□
DRUG_EXPOSURE_START_DATETIME	2019-05-01 00:00:00	□□□□□□□□0□□□□□
DRUG_EXPOSURE_END_DATE	2019-05-31	□□□ + □□□□□□□□□
DRUG_EXPOSURE_END_DATETIME	2019-05-31 00:00:00	□□□□□□□□0□□□□□
VERBATIM_END_DATE	NULL	□□□□□□□□□
DRUG_TYPE_CONCEPT_ID	38000177	38000177 □□□□□□□□□□□□□□
STOP_REASON	NULL	
REFILLS	NULL	
QUANTITY	NULL	□□□□□□□□□
DAYS_SUPPLY	30	□□□□□□□□□□□□□
SIG	NULL	□□□□□□□□□
ROUTE_CONCEPT_ID	4132161	4132161 □□□□□□□□□
LOT_NUMBER	NULL	□□□□□□□□□
PROVIDER_ID	NULL	□□□□□□□□□
VISIT_OCCURRENCE_ID	NULL	□□□□□□□□□□□□□□□□□□□□□□□□
VISIT_DETAIL_ID	NULL	
DRUG_SOURCE_VALUE	76168009520	□□□□□NDC□□□□
DRUG_SOURCE_CONCEPT_ID	583945	583945 □□□□□□□□□□□□□□□□ (NDC□□□□76168009520□)□
ROUTE_SOURCE_VALUE	NULL	

□□ 4.4

□□□□□□□□□□□□□□□□CONDITION_OCCURRENCE□□□□□□□□□□□□□□□□□□□□

```
library(DatabaseConnector)
connection <- connect(connectionDetails)
sql <- "SELECT *
FROM @cdm.condition_occurrence
WHERE condition_concept_id = 192671;"

result <- renderTranslateQuerySql(connection, sql, cdm = "main")
head(result)
```

```
##   CONDITION_OCCURRENCE_ID PERSON_ID CONDITION_CONCEPT_ID ...
```

```
## 1                    4657       273                192671 ...
## 2                    1021        61                192671 ...
## 3                    5978       351                192671 ...
## 4                    9798       579                192671 ...
## 5                    9301       549                192671 ...
## 6                    1997       116                192671 ...
```

□□ 4.5

□□□□□□□□□□□□□□□□CONDITION_OCCURRENCE□□□□□□CONDITION_SOURCE_VALUE

```r
sql <- "SELECT *
FROM @cdm.condition_occurrence
WHERE condition_source_value = 'K92.2';"

result <- renderTranslateQuerySql(connection, sql, cdm = "main")
head(result)
```

```
##   CONDITION_OCCURRENCE_ID PERSON_ID CONDITION_CONCEPT_ID ...
## 1                    4657       273                192671 ...
## 2                    1021        61                192671 ...
## 3                    5978       351                192671 ...
## 4                    9798       579                192671 ...
## 5                    9301       549                192671 ...
## 6                    1997       116                192671 ...
```

□□ 4.6

□□□□□OBSERVATION_PERIOD□□□□□□□□□□□□□□□

```r
library(DatabaseConnector)
connection <- connect(connectionDetails)
sql <- "SELECT *
FROM @cdm.observation_period
WHERE person_id = 61;"

renderTranslateQuerySql(connection, sql, cdm = "main")
```

```
##   OBSERVATION_PERIOD_ID PERSON_ID OBSERVATION_PERIOD_START_DATE ...
## 1                    61        61                     1968-01-21 ...
```

E.2

□□ 5.1

□□□□□ ID 192671 (“□□□□□”)

□□ 5.2

ICD-10CM□□□□

- K29.91 “□□□□□□□□□□□□□□□□□”
- K92.2 “□□□□□□□□□□□”

ICD-9CM□□□□

- 578 “□□□□□”
- 578.9 “□□□□□□□□□□□”

□□ 5.3

MedDRA□□□□Preferred Terms, PT)□

- “□□□□□” (□□□□□ID 35707864)
- “□□□” (□□□□□ID 35707858)

E.3 ETL (Extract-Transform-Load)

□□ 6.1

A) □□□□□□□□CDM□□□□□□□□□ETL□□□□□□□□□
B) □□□□□□□□□□□□□□□□□□□□□□□□□
C) □□□□□□ETL□□□□□□□
D) □□□□□□□□□□□□□□□

□□ 6.2

□□□	□	□□
PERSON_ID	A123B456	□□□□□□□□□□□□□□□□□□□□□□
GENDER_CONCEPT_ID	8532	
YEAR_OF_BIRTH	NULL	□□□□□□□□□□□□□□□□□□□□□□
MONTH_OF_BIRTH	NULL	
DAY_OF_BIRTH	NULL	
RACE_CONCEPT_ID	0	□□□WHITE□□□□□□□8527□□□
ETHNICITY_CONCEPT_ID	8527	□□□□□□□□□□□□□□□□□□□□□□
PERSON_SOURCE_VALUE	A123B456	
GENDER_SOURCE_VALUE	F	
RACE_SOURCE_VALUE	WHITE	
ETHNICITY_SOURCE_VALUE	NONE PROVIDED	

□□ **6.3**

□□□	□
VISIT_OCCURRENCE_ID	1
PERSON_ID	11
VISIT_START_DATE	2004-09-26
VISIT_END_DATE	2004-09-30
VISIT_CONCEPT_ID	9201
VISIT_SOURCE_VALUE	inpatient

E.4

□□ **7.1**

1. □□□□□
2. □□□□□□□□
3. □□□□□□□□

□□ **7.2**

□□□□□□□

```
library(DatabaseConnector)
connection <- connect(connectionDetails)
sql <- "SELECT COUNT(DISTINCT(person_id)) AS person_count
FROM @cdm.drug_exposure
INNER JOIN @cdm.concept_ancestor
  ON drug_concept_id = descendant_concept_id
INNER JOIN @cdm.concept ingredient
  ON ancestor_concept_id = ingredient.concept_id
WHERE LOWER(ingredient.concept_name) = 'celecoxib'
  AND ingredient.concept_class_id = 'Ingredient'
  AND ingredient.standard_concept = 'S';"

renderTranslateQuerySql(connection, sql, cdm = "main")
```

```
##   PERSON_COUNT
## 1         1844
```

COUNT(DISTINCT(person_id))□□□

```
INNER JOIN @cdm.condition_occurrence
  ON condition_start_date >= drug_era_start_date
    AND condition_start_date <= drug_era_end_date
INNER JOIN @cdm.concept_ancestor
  ON condition_concept_id =descendant_concept_id
WHERE LOWER(ingredient.concept_name) = 'celecoxib'
  AND ingredient.concept_class_id = 'Ingredient'
  AND ingredient.standard_concept = 'S'
  AND ancestor_concept_id = 192671;"

renderTranslateQuerySql(connection, sql, cdm = "main")
```

```
##   DIAGNOSE_COUNT
## 1             41
```

□□□□□DRUG_EXPOSURE□□□□□□□□□DRUG_ERA□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

E.6

□□ 10.1

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□:

- □□□□□□□□□□□□□□□
- □□□16□□□
- □□□□□□□□□□365□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□

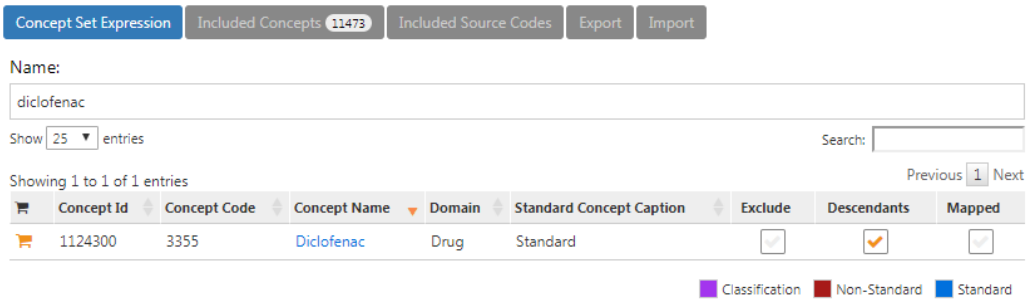


Figure E.2: □□□□□□□□□□□□□□□□□□

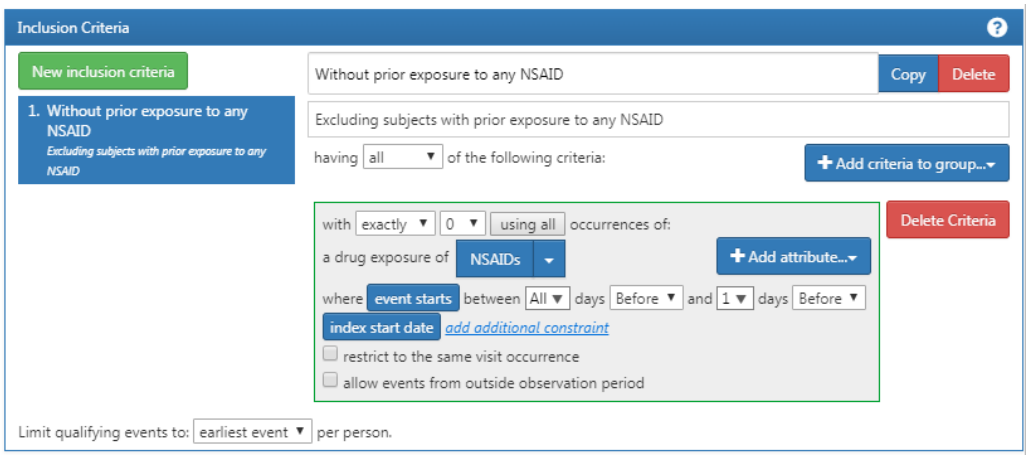


Figure E.3: NSAID□□□□□□□□□□□□□□□□□

NSAIDs□□□□□□□□□□□□NSAIDs□□□□□□□□□□□□

Concept Set Expression | Included Concepts 4401 | Included Source Codes | Export | Import

Name:

Broad malignancies

Show 25 entries Search:

Showing 1 to 1 of 1 entries Previous 1 Next

	Concept Id	Concept Code	Concept Name	Domain	Standard Concept Caption	Exclude	Descendants	Mapped
	443392	363346000	Malignant neoplastic disease	Condition	Standard		✓	

Classification Non-Standard Standard

Figure E.6: □□□□□□□□□□□□□□□□□□

Cohort Exit

Event Persistence:

Event will persist until: end of a continuous drug exposure

Continuous Exposure Persistence:

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

Concept set containing the drug(s) of interest: diclofenac

- Persistence window: allow for a maximum of 30 days between exposure records when inferring the era of persistence exposure
- Surveillance window: add 0 days to the end of the era of persistence exposure as an additional period of surveillance prior to cohort exit.

Censoring Events:

Exit Cohort based on the following criteria: + Add Censoring Event

No censoring events selected.

Figure E.7: □□□□□□□□□□□□□□□□□

```
INTO #diagnoses
FROM @cdm.condition_occurrence
WHERE condition_concept_id IN (
    SELECT descendant_concept_id
    FROM @cdm.concept_ancestor
    WHERE ancestor_concept_id = 4329847 --
)
  AND condition_concept_id NOT IN (
    SELECT descendant_concept_id
    FROM @cdm.concept_ancestor
    WHERE ancestor_concept_id = 314666 --
);"

renderTranslateExecuteSql(connection, sql, cdm = "main")
```

□□□

E.7

□□ 11.1

ATLAS□ Data Sources□□□□□

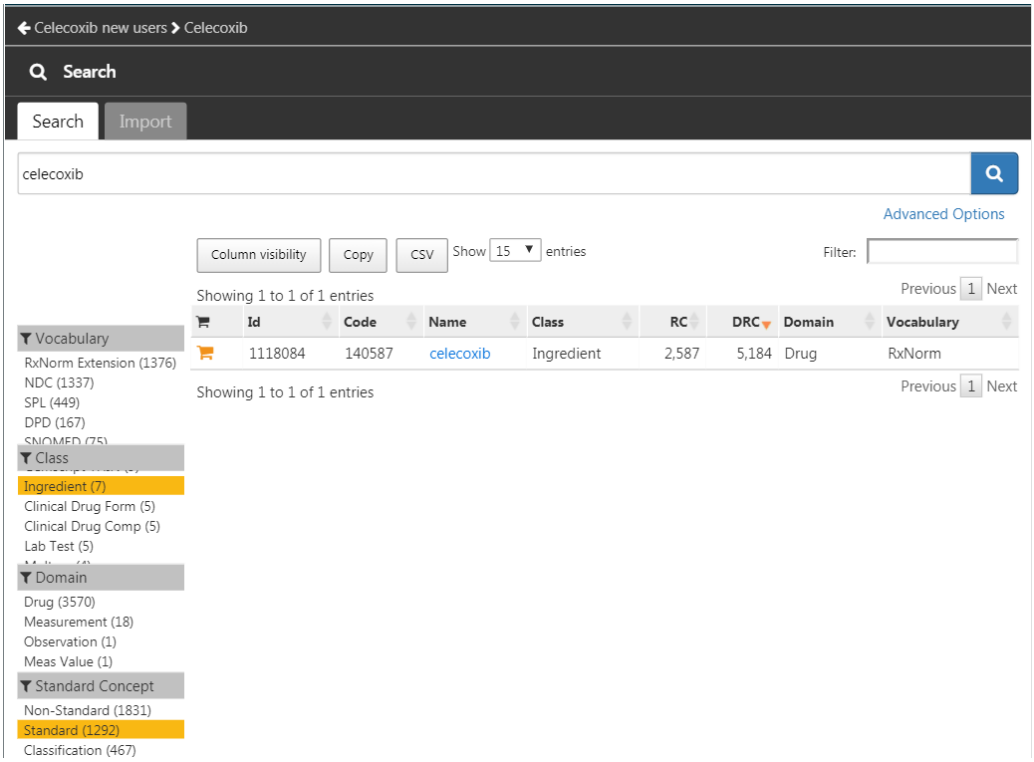


Figure E.9: □□□Celecoxib (□□□□□□)□□□□□□□□□□□□□□

attribute… (□□□□□…)□□□□□□□□□□Add First Exposure Criteria (□□□□□□□□□□□)□□□□
E.10 □□□□□□□□□□□□

New Characterization

Celecoxib new users characterization

Design | Executions | Utilities

Cohort characterization is defined as the process of generating cohort level descriptive summary statistics from person level covariate data. Summary statistics of these person level covariates may be count, mean, sd, var, min, max, median, range, and quantiles. In addition, covariates during a period may be stratified into temporal units of time for time-series analysis such as fixed intervals of time relative to cohort_start_date (e.g. every 7 days, every 30 days etc.), or in absolute calendar intervals such as calendar-week, calendar-month, calendar-quarter, calendar-year.

Cohort definitions

Import

Show 10 entries Search:

ID	Name		
1771701	Celecoxib new users	Edit cohort	Remove

Showing 1 to 1 of 1 entries Previous 1 Next

Feature analyses

Import

Show 10 entries Search:

ID	Name	Description	Actions
15	Drug Group Era Any Time Prior	One covariate per drug rolled up to ATC groups in the drug_era table overlapping with any time prior to index.	Remove
27	Condition Group Era Any Time Prior	One covariate per condition era rolled up to groups in the condition_era table overlapping with any time prior to index.	Remove

Showing 1 to 2 of 2 entries Previous 1 Next

Figure E.11: □□□□□□

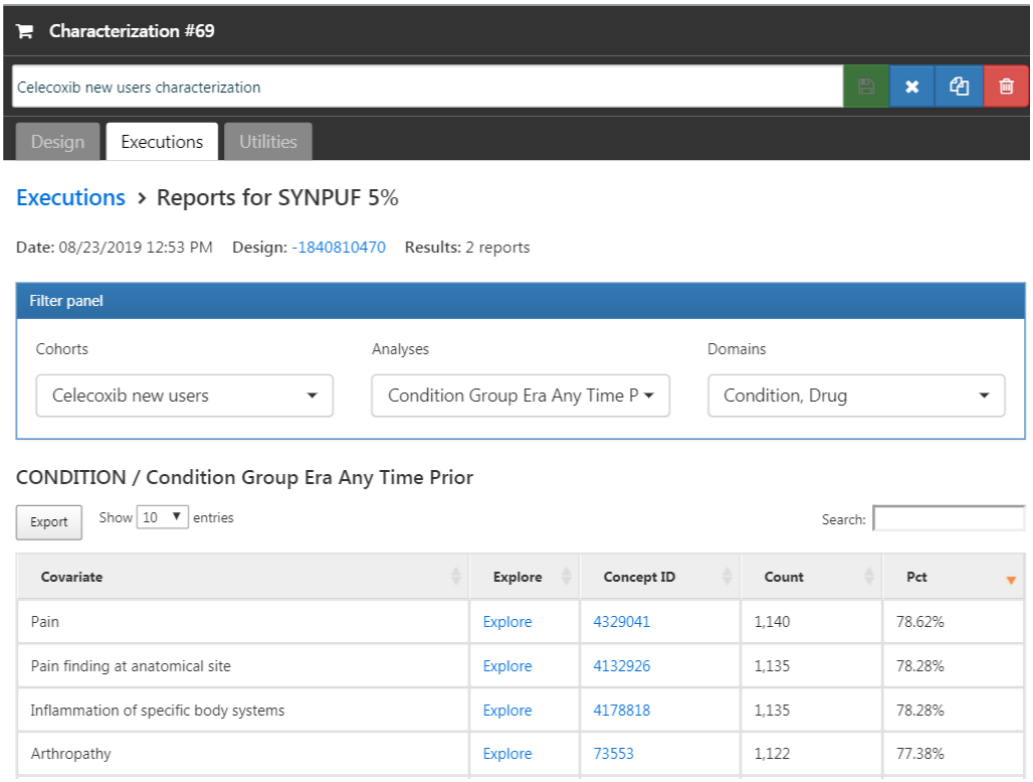


Figure E.12: □□□□□□□

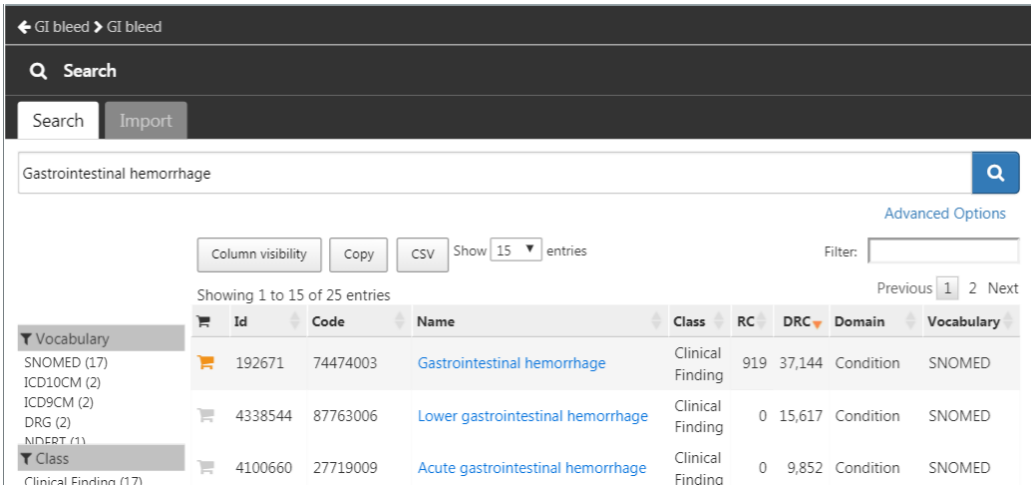


Figure E.13: □Gastrointestinal hemorrhage (□□□□□)□□□□□□□□□□□□□

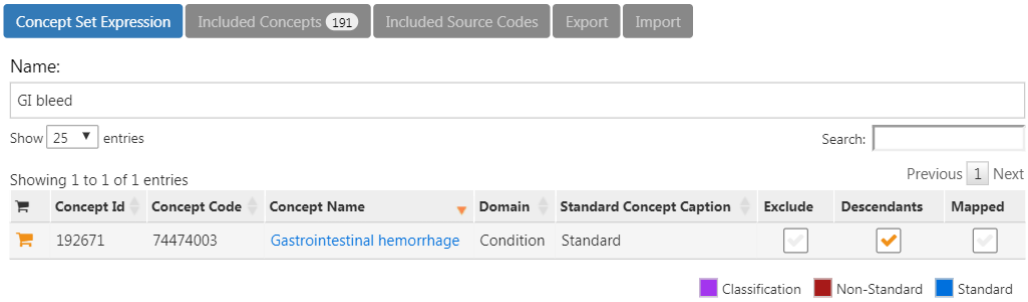


Figure E.14: □Gastrointestinal hemorrhage (□□□□□)□□□□□□□□□□□□□

□Definition (□□)□□□□□□□□□+Add Initial Event (□□□□□□□□□)□□□□□□□□□□□Add Condition Occurrence (□□□□□□□□□□□□□)□□□□□□□□□□□□□

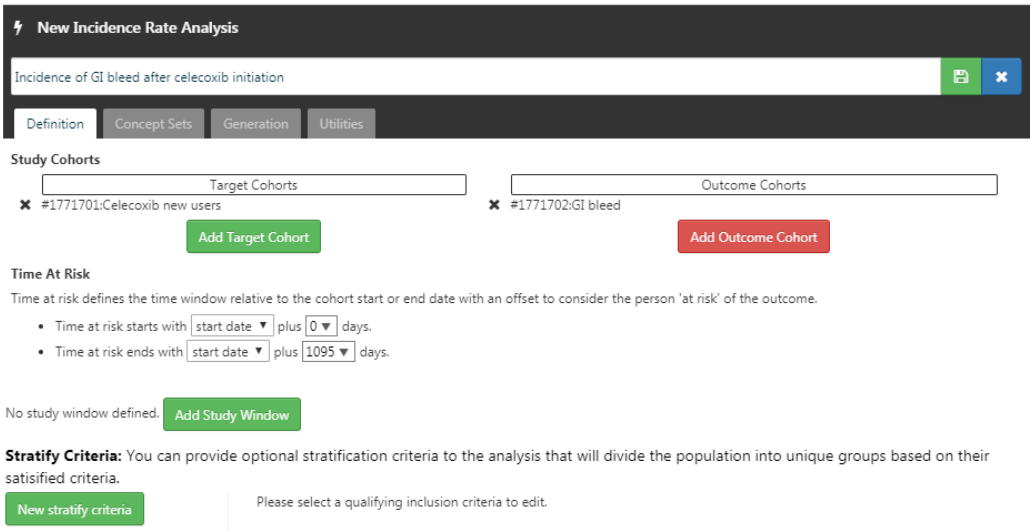


Figure E.16: □□□□□□

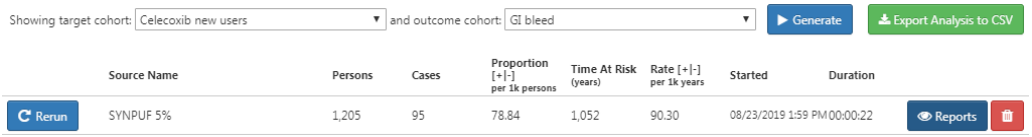


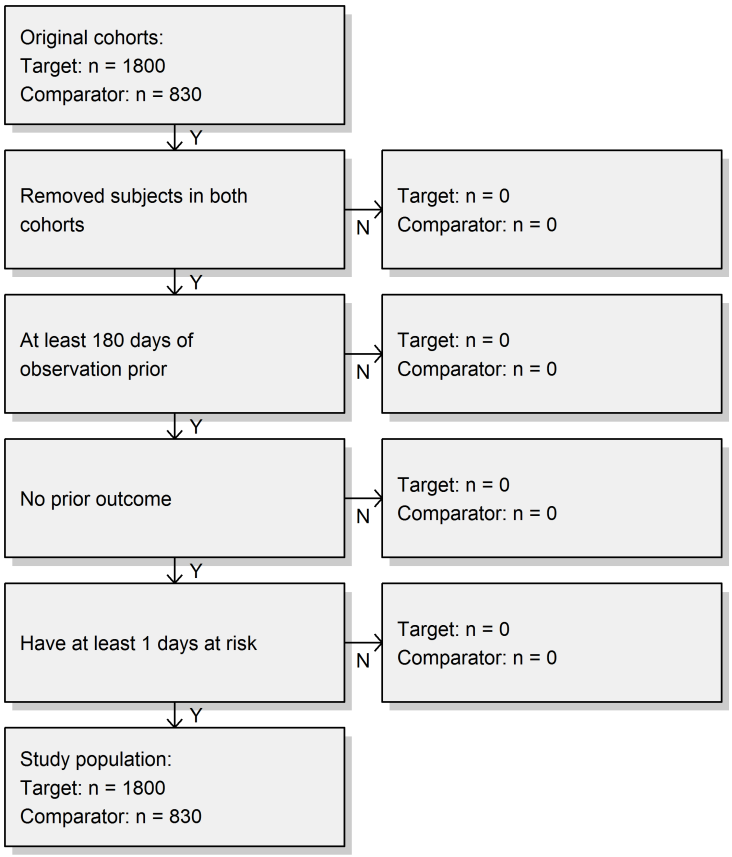
Figure E.17: □□□□□□

E.8

□□ 12.1

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□2□□□□□□□□□□□□□□□□□□□

```
studyPop <- createStudyPopulation(
  cohortMethodData = cmData,
  outcomeId = 3,
  washoutPeriod = 180,
  removeDuplicateSubjects = "remove all",
  removeSubjectsWithPriorOutcome = TRUE,
  riskWindowStart = 0,
  startAnchor = "cohort start",
  riskWindowEnd = 99999)
drawAttritionDiagram(studyPop)
```



□□□□□□□□□

```
model <- fitOutcomeModel(population = studyPop,
                         modelType = "cox")
model
```

```
##     cox
##   FALSE
##     FALSE
##          FALSE
##    OK
##
##                .95     .95
##     1.34612   1.10065   1.65741        0.29723                    0.1044
```

□□□

□□ 12.5

□□□□□□□□□

```
##     cox
##   TRUE
##     FALSE
##     (IPTW)  FALSE
##    OK
##
##                .95    .95
##     1.13211   0.92132   1.40008         0.12409                  0.1068
```

□□□□□□□□□□□□□□□□□□□□□□□□□□95□□□□□□□1□□□□□□□□□□□□□□□□□□□□□□□□□□2□□

E.9

□□ **13.1**

□□□getPlpData□□□□□□□□□□

```r
library(PatientLevelPrediction)
covSettings <- createCovariateSettings(
  useDemographicsGender = TRUE,
  useDemographicsAge = TRUE,
  useConditionGroupEraLongTerm = TRUE,
  useConditionGroupEraAnyTimePrior = TRUE,
  useDrugGroupEraLongTerm = TRUE,
  useDrugGroupEraAnyTimePrior = TRUE,
  useVisitConceptCountLongTerm = TRUE,
  longTermStartDays = -365,
  endDays = -1)

plpData <- getPlpData(connectionDetails = connectionDetails,
                      cdmDatabaseSchema = "main",
                      cohortDatabaseSchema = "main",
                      cohortTable = "cohort",
                      cohortId = 4,
                      covariateSettings = covSettings,
                      outcomeDatabaseSchema = "main",
                      outcomeTable = "cohort",
                      outcomeIds = 3)

summary(plpData)
```

```
## plpData
##
## At risk        ID -1
##       ID 3
##
##   2630
```

```
##
##
##
## 3           479     479
##
##
##     245
##       54079
```

□□ 13.2

□□□□□□□□□

Shiny□□□□□□□□□□□□□□□□□□□□□□□□□□

```
viewPlp(lassoResults)
```

□□□□□□□□E.18□□□□□□□□□□□□□□□□□□□

```
  resultsDatabaseSchema = "main",
  cdmSourceName = "Eunomia",
  outputFolder = "C:/dataQualityExample")
```

□□ 15.3

□□□□□□□□□□□□□□□□□□□□:

```
DataQualityDashboard::viewDqDashboard(
  "C:/dataQualityExample/Eunomia/results_Eunomia.json")
```

E.11

Bibliography

Allison, D. B., Brown, A. W., George, B. J., and Kaiser, K. A. (2016). Reproducibility: A tragedy of errors. *Nature*, 530(7588):27–29.

Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical statistics*, 10(2):150–161.

Banda, J. M., Halpern, Y., Sontag, D., and Shah, N. H. (2017). Electronic phenotyping with APHRODITE and the Observational Health Sciences and Informatics (OHDSI) data network. *AMIA Jt Summits Transl Sci Proc*, 2017:48–57.

Boland, M. R., Parhi, P., Li, L., Miotto, R., Carroll, R., Iqbal, U., Nguyen, P. A., Schuemie, M., You, S. C., Smith, D., Mooney, S., Ryan, P., Li, Y. J., Park, R. W., Denny, J., Dudley, J. T., Hripcsak, G., Gentine, P., and Tatonetti, N. P. (2017). Uncovering exposures responsible for birth season - disease effects: a global study. *J Am Med Inform Assoc*.

Byrd, J. B., Adam, A., and Brown, N. J. (2006). Angiotensin-converting enzyme inhibitor-associated angioedema. *Immunol Allergy Clin North Am*, 26(4):725–737.

Cepeda, M. S., Reps, J., Fife, D., Blacketer, C., Stang, P., and Ryan, P. (2018). Finding treatment-resistant depression in real-world data: How a data-driven approach compares with expert-based heuristics. *Depress Anxiety*, 35(3):220–228.

Chen, X., Dallmeier-Tiessen, S., Dasler, R., Feger, S., Fokianos, P., Gonzalez, J. B., Hirvonsalo, H., Kousidis, D., Lavasa, A., Mele, S., Rodriguez, D. R., Šimko, T., Smith, T., Trisovic, A., Trzcinska, A., Tsanaktsidis, I., Zimmermann, M., Cranmer, K., Heinrich, L., Watts, G., Hildreth, M., Iglesias, L. L., Lassila-Perini, K., and Neubert, S. (2018). Open is not enough. *Nature Physics*, 15(2):113–119.

Cicardi, M., Zingale, L. C., Bergamaschini, L., and Agostoni, A. (2004). Angioedema associated with angiotensin-converting enzyme inhibitor use: outcome after switching to a different treatment. *Arch. Intern. Med.*, 164(8):910–913.

Dasu, T. and Johnson, T. (2003). *Exploratory data mining and data cleaning*, volume 479. John Wiley & Sons.

Defalco, F. J., Ryan, P. B., and Soledad Cepeda, M. (2013). Applying standardized drug

terminologies to observational healthcare databases: a case study on opioid exposure. *Health Serv Outcomes Res Methodol*, 13(1):58–67.

Duke, J. D., Ryan, P. B., Suchard, M. A., Hripcsak, G., Jin, P., Reich, C., Schwalm, M. S., Khoma, Y., Wu, Y., Xu, H., Shah, N. H., Banda, J. M., and Schuemie, M. J. (2017). Risk of angioedema associated with levetiracetam compared with phenytoin: Findings of the observational health data sciences and informatics research network. *Epilepsia*, 58(8):e101–e106.

Farrington, C. P. (1995). Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics*, 51(1):228–235.

Farrington, C. P., Anaya-Izquierdo, K., Whitaker, H. J., Hocine, M. N., Douglas, I., and Smeeth, L. (2011). Self-controlled case series analysis with event-dependent observation periods. *Journal of the American Statistical Association*, 106(494):417–426.

Fuller, W. A. (2009). *Measurement error models*, volume 305. John Wiley & Sons.

Garza, M., Del Fiol, G., Tenenbaum, J., Walden, A., and Zozus, M. N. (2016). Evaluating common data models for use with a longitudinal community registry. *J Biomed Inform*, 64:333–341.

Hernan, M. A., Hernandez-Diaz, S., Werler, M. M., and Mitchell, A. A. (2002). Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am. J. Epidemiol.*, 155(2):176–184.

Hernan, M. A. and Robins, J. M. (2016). Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am. J. Epidemiol.*, 183(8):758–764.

Hripcsak, G., Duke, J. D., Shah, N. H., Reich, C. G., Huser, V., Schuemie, M. J., Suchard, M. A., Park, R. W., Wong, I. C. K., Rijnbeek, P. R., van der Lei, J., Pratt, N., Norén, G. N., Li, Y.-C., Stang, P. E., Madigan, D., and Ryan, P. B. (2015). Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Studies in health technology and informatics*, 216:574–578.

Hripcsak, G., Levine, M. E., Shang, N., and Ryan, P. B. (2018). Effect of vocabulary mapping for conditions on phenotype cohorts. *J Am Med Inform Assoc*, 25(12):1618–1625.

Hripcsak, G., Ryan, P. B., Duke, J. D., Shah, N. H., Park, R. W., Huser, V., Suchard, M. A., Schuemie, M. J., DeFalco, F. J., Perotte, A., Banda, J. M., Reich, C. G., Schilling, L. M., Matheny, M. E., Meeker, D., Pratt, N., and Madigan, D. (2016). Characterizing treatment pathways at scale using the OHDSI network. *Proceedings of the National Academy of Sciences*, 113(27):7329–7336.

Hripcsak, G., Shang, N., Peissig, P. L., Rasmussen, L. V., Liu, C., Benoit, B., Carroll, R. J., Carrell, D. S., Denny, J. C., Dikilitas, O., Gainer, V. S., Marie Howell, K., Klann, J. G., Kullo, I. J., Lingren, T., Mentch, F. D., Murphy, S. N., Natarajan, K., Pacheco, J. A., Wei, W. Q., Wiley, K., and Weng, C. (2019). Facilitating phenotype transfer using a common data model. *J Biomed Inform*, page 103253.

Johnston, S. S., Morton, J. M., Kalsekar, I., Ammann, E. M., Hsiao, C. W., and Reps, J. (2019). Using Machine Learning Applied to Real-World Healthcare Data for Predictive Analytics: An Applied Example in Bariatric Surgery. *Value Health*, 22(5):580–586.

Kahn, M. G., Callahan, T. J., Barnard, J., Bauck, A. E., Brown, J., Davidson, B. N., Estiri, H., Goerg, C., Holve, E., Johnson, S. G., Liaw, S.-T., Hamilton-Lopez, M., Meeker, D., Ong, T. C., Ryan, P. B., Shang, N., Weiskopf, N. G., Weng, C., Zozus, M. N., and Schilling, L. (2016). A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *EGEMS (Washington, DC)*, 4(1):1244.

Maclure, M. (1991). The case-crossover design: a method for studying transient effects on the risk of acute events. *Am. J. Epidemiol.*, 133(2):144–153.

Madigan, D., Ryan, P. B., and Schuemie, M. (2013a). Does design matter? Systematic evaluation of the impact of analytical choices on effect estimates in observational studies. *Ther Adv Drug Saf*, 4(2):53–62.

Madigan, D., Ryan, P. B., Schuemie, M., Stang, P. E., Overhage, J. M., Hartzema, A. G., Suchard, M. A., DuMouchel, W., and Berlin, J. A. (2013b). Evaluating the impact of database heterogeneity on observational study results. *Am. J. Epidemiol.*, 178(4):645–651.

Magid, D. J., Shetterly, S. M., Margolis, K. L., Tavel, H. M., O'Connor, P. J., Selby, J. V., and Ho, P. M. (2010). Comparative effectiveness of angiotensin-converting enzyme inhibitors versus beta-blockers as second-line therapy for hypertension. *Circ Cardiovasc Qual Outcomes*, 3(5):453–458.

Makadia, R. and Ryan, P. B. (2014). Transforming the Premier Perspective Hospital Database into the Observational Medical Outcomes Partnership (OMOP) Common Data Model. *EGEMS (Wash DC)*, 2(1):1110.

Matcho, A., Ryan, P., Fife, D., and Reich, C. (2014). Fidelity assessment of a clinical practice research datalink conversion to the OMOP common data model. *Drug Saf*, 37(11):945–959.

Norman, J. L., Holmes, W. L., Bell, W. A., and Finks, S. W. (2013). Life-threatening ACE inhibitor-induced angioedema after eleven years on lisinopril. *J Pharm Pract*, 26(4):382–388.

Oliveira, J. L., Trifan, A., and Silva, L. A. B. (2019). EMIF catalogue: A collaborative platform for sharing and reusing biomedical data. *International Journal of Medical Informatics*, 126:35–45.

Olsen, L., Aisner, D., McGinnis, J. M., et al. (2007). *The learning healthcare system: workshop summary*. Natl Academy Pr.

O'Mara, N. B. and O'Mara, E. M. (1996). Delayed onset of angioedema with angiotensin-converting enzyme inhibitors: case report and review of the literature. *Pharmacotherapy*, 16(4):675–679.

Overhage, J. M., Ryan, P. B., Reich, C. G., Hartzema, A. G., and Stang, P. E. (2012). Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc*, 19(1):54–60.

Perkins, N. J., Cole, S. R., Harel, O., Tchetgen Tchetgen, E. J., Sun, B., Mitchell, E. M., and Schisterman, E. F. (2017). Principled approaches to missing data in epidemiologic studies. *American journal of epidemiology*, 187(3):568–575.

Powers, B. J., Coeytaux, R. R., Dolor, R. J., Hasselblad, V., Patel, U. D., Yancy, W. S., Gray, R. N., Irvine, R. J., Kendrick, A. S., and Sanders, G. D. (2012). Updated report on comparative effectiveness of ACE inhibitors, ARBs, and direct renin inhibitors for patients with essential hypertension: much more data, little new information. *J Gen Intern Med*, 27(6):716–729.

Ramcharran, D., Qiu, H., Schuemie, M. J., and Ryan, P. B. (2017). Atypical Antipsychotics and the Risk of Falls and Fractures Among Older Adults: An Emulation Analysis and an Evaluation of Additional Confounding Control Strategies. *J Clin Psychopharmacol*, 37(2):162–168.

Rassen, J. A., Shelat, A. A., Myers, J., Glynn, R. J., Rothman, K. J., and Schneeweiss, S. (2012). One-to-many propensity score matching in cohort studies. *Pharmacoepidemiol Drug Saf*, 21 Suppl 2:69–80.

Reps, J. M., Rijnbeek, P. R., and Ryan, P. B. (2019). Identifying the DEAD: Development and Validation of a Patient-Level Model to Predict Death Status in Population-Level Claims Data. *Drug Saf*.

Reps, J. M., Schuemie, M. J., Suchard, M. A., Ryan, P. B., and Rijnbeek, P. R. (2018). Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. *Journal of the American Medical Informatics Association*, 25(8):969–975.

Roebuck, K. (2012). *Data quality: high-impact strategies-what you need to know: definitions, adoptions, impact, benefits, maturity, vendors*. Emereo Publishing.

Rosenbaum, P. and Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2(3-4):169–188.

Ryan, P. B., Buse, J. B., Schuemie, M. J., DeFalco, F., Yuan, Z., Stang, P. E., Berlin, J. A., and Rosenthal, N. (2018). Comparative effectiveness of canagliflozin, SGLT2 inhibitors and non-SGLT2 inhibitors on the risk of hospitalization for heart failure and amputation in patients with type 2 diabetes mellitus: A real-world meta-analysis of 4 observational databases (OBSERVE-4D). *Diabetes Obes Metab*, 20(11):2585–2597.

Ryan, P. B., Madigan, D., Stang, P. E., Overhage, J. M., Racoosin, J. A., and Hartzema, A. G. (2012). Empirical assessment of methods for risk identification in healthcare data:

results from the experiments of the Observational Medical Outcomes Partnership. *Stat Med*, 31(30):4401–4415.

Ryan, P. B., Schuemie, M. J., and Madigan, D. (2013a). Empirical performance of a self-controlled cohort method: lessons for developing a risk identification and analysis system. *Drug Saf*, 36 Suppl 1:95–106.

Ryan, P. B., Schuemie, M. J., Ramcharran, D., and Stang, P. E. (2017). Atypical Antipsychotics and the Risks of Acute Kidney Injury and Related Outcomes Among Older Adults: A Replication Analysis and an Evaluation of Adapted Confounding Control Strategies. *Drugs Aging*, 34(3):211–219.

Ryan, P. B., Stang, P. E., Overhage, J. M., Suchard, M. A., Hartzema, A. G., DuMouchel, W., Reich, C. G., Schuemie, M. J., and Madigan, D. (2013b). A comparison of the empirical performance of methods for a risk identification system. *Drug Saf*, 36 Suppl 1:S143–158.

Sabroe, R. A. and Black, A. K. (1997). Angiotensin-converting enzyme (ACE) inhibitors and angio-oedema. *Br. J. Dermatol.*, 136(2):153–158.

Schneeweiss, S. (2018). Automated data-adaptive analytics for electronic healthcare data to study causal treatment effects. *Clin Epidemiol*, 10:771–788.

Schuemie, M. J., Hripcsak, G., Ryan, P. B., Madigan, D., and Suchard, M. A. (2016). Robust empirical calibration of p-values using observational data. *Stat Med*, 35(22):3883–3888.

Schuemie, M. J., Hripcsak, G., Ryan, P. B., Madigan, D., and Suchard, M. A. (2018a). Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proc. Natl. Acad. Sci. U.S.A.*, 115(11):2571–2577.

Schuemie, M. J., Ryan, P. B., DuMouchel, W., Suchard, M. A., and Madigan, D. (2014). Interpreting observational studies: why empirical calibration is needed to correct p-values. *Stat Med*, 33(2):209–218.

Schuemie, M. J., Ryan, P. B., Hripcsak, G., Madigan, D., and Suchard, M. A. (2018b). Improving reproducibility by using high-throughput observational studies with empirical calibration. *Philos Trans A Math Phys Eng Sci*, 376(2128).

Simpson, S. E., Madigan, D., Zorych, I., Schuemie, M. J., Ryan, P. B., and Suchard, M. A. (2013). Multiple self-controlled case series for large-scale longitudinal observational databases. *Biometrics*, 69(4):893–902.

Slater, E. E., Merrill, D. D., Guess, H. A., Roylance, P. J., Cooper, W. D., Inman, W. H. W., and Ewan, P. W. (1988). Clinical Profile of Angioedema Associated With Angiotensin Converting-Enzyme Inhibition. *JAMA*, 260(7):967–970.

Stang, P. E., Ryan, P. B., Racoosin, J. A., Overhage, J. M., Hartzema, A. G., Reich, C., Welebob, E., Scarnecchia, T., and Woodcock, J. (2010). Advancing the science

for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. *Ann. Intern. Med.*, 153(9):600–606.

Suchard, M. A., Simpson, S. E., Zorych, I., Ryan, P. B., and Madigan, D. (2013). Massive parallelization of serial inference algorithms for a complex generalized linear model. *ACM Trans. Model. Comput. Simul.*, 23(1):10:1–10:17.

Suissa, S. (1995). The case-time-control design. *Epidemiology*, 6(3):248–253.

Swerdel, J. N., Hripcsak, G., and Ryan, P. B. (2019). PheValuator: Development and Evaluation of a Phenotype Algorithm Evaluator. *J Biomed Inform*, page 103258.

Thompson, T. and Frable, M. A. (1993). Drug-induced, life-threatening angioedema revisited. *Laryngoscope*, 103(1 Pt 1):10–12.

Tian, Y., Schuemie, M. J., and Suchard, M. A. (2018). Evaluating large-scale propensity score performance through real-world and synthetic data experiments. *Int J Epidemiol*, 47(6):2005–2014.

Toh, S., Reichman, M. E., Houstoun, M., Ross Southworth, M., Ding, X., Hernandez, A. F., Levenson, M., Li, L., McCloskey, C., Shoaibi, A., Wu, E., Zornberg, G., and Hennessy, S. (2012). Comparative risk for angioedema associated with the use of drugs that target the renin-angiotensin-aldosterone system. *Arch. Intern. Med.*, 172(20):1582–1589.

van der Lei, J. (1991). Use and abuse of computer-stored medical records. *Methods of information in medicine*, 30(02):79–80.

Vandenbroucke, J. P. and Pearce, N. (2012). Case-control studies: basic concepts. *Int J Epidemiol*, 41(5):1480–1489.

Vashisht, R., Jung, K., Schuler, A., Banda, J. M., Park, R. W., Jin, S., Li, L., Dudley, J. T., Johnson, K. W., Shervey, M. M., Xu, H., Wu, Y., Natrajan, K., Hripcsak, G., Jin, P., Van Zandt, M., Reckard, A., Reich, C. G., Weaver, J., Schuemie, M. J., Ryan, P. B., Callahan, A., and Shah, N. H. (2018). Association of Hemoglobin A1c Levels With Use of Sulfonylureas, Dipeptidyl Peptidase 4 Inhibitors, and Thiazolidinediones in Patients With Type 2 Diabetes Treated With Metformin: Analysis From the Observational Health Data Sciences and Informatics Initiative. *JAMA Netw Open*, 1(4):e181755.

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., and Vandenbroucke, J. P. (2008). The strengthening the reporting of observational studies in epidemiology (strobe) statement: guidelines for reporting observational studies. *Journal of Clinical Epidemiology*, 61(4):344 – 349.

Voss, E. A., Ma, Q., and Ryan, P. B. (2015a). The impact of standardizing the definition of visits on the consistency of multi-database observational health research. *BMC Med Res Methodol*, 15:13.

Voss, E. A., Makadia, R., Matcho, A., Ma, Q., Knoll, C., Schuemie, M., DeFalco, F. J., Londhe, A., Zhu, V., and Ryan, P. B. (2015b). Feasibility and utility of applications

of the common data model to multiple, disparate observational health databases. *J Am Med Inform Assoc*, 22(3):553–564.

Walker, A. M., Patrick, A. R., Lauer, M. S., Hornbrook, M. C., Marin, M. G., Platt, R., Roger, V. L., Stang, P., and Schneeweiss, S. (2013). A tool for assessing the feasibility of comparative effectiveness research. *Comp Eff Res*, 3:11–20.

Wang, Y., Desai, M., Ryan, P. B., DeFalco, F. J., Schuemie, M. J., Stang, P. E., Berlin, J. A., and Yuan, Z. (2017). Incidence of diabetic ketoacidosis among patients with type 2 diabetes mellitus treated with SGLT2 inhibitors and other antihyperglycemic agents. *Diabetes Res. Clin. Pract.*, 128:83–90.

Weinstein, R. B., Ryan, P., Berlin, J. A., Matcho, A., Schuemie, M., Swerdel, J., Patel, K., and Fife, D. (2017). Channeling in the Use of Nonprescription Paracetamol and Ibuprofen in an Electronic Medical Records Database: Evidence and Implications. *Drug Saf*, 40(12):1279–1292.

Weiskopf, N. G. and Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association: JAMIA*, 20(1):144–151.

Whelton, P. K., Carey, R. M., Aronow, W. S., Casey, D. E., Collins, K. J., Dennison Himmelfarb, C., DePalma, S. M., Gidding, S., Jamerson, K. A., Jones, D. W., MacLaughlin, E. J., Muntner, P., Ovbiagele, B., Smith, S. C., Spencer, C. C., Stafford, R. S., Taler, S. J., Thomas, R. J., Williams, K. A., Williamson, J. D., and Wright, J. T. (2018). 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*, 138(17):e426–e483.

Whitaker, H. J., Farrington, C. P., Spiessens, B., and Musonda, P. (2006). Tutorial in biostatistics: the self-controlled case series method. *Stat Med*, 25(10):1768–1797.

Who, A. (2013). Global brief on hypertension. *World Health Organization*.

Wickham, H. (2015). *R Packages*. O'Reilly Media, Inc., 1st edition.

Wikipedia (2019a). Open science — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Open%20science&oldid=900178688. [Online; accessed 24-June-2019].

Wikipedia (2019b). Science 2.0 — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Science%202.0&oldid=887565958. [Online; accessed 09-July-2019].

Wikiquote (2019). Ronald fisher — wikiquote,. [Online; accessed 2-August-2019].

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman,

J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S. A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, 3:160018.

Yuan, Z., DeFalco, F. J., Ryan, P. B., Schuemie, M. J., Stang, P. E., Berlin, J. A., Desai, M., and Rosenthal, N. (2018). Risk of lower extremity amputations in people with type 2 diabetes mellitus treated with sodium-glucose co-transporter-2 inhibitors in the USA: A retrospective cohort study. *Diabetes Obes Metab*, 20(3):582–589.

Zaman, M. A., Oparil, S., and Calhoun, D. A. (2002). Drugs targeting the renin-angiotensin-aldosterone system. *Nat Rev Drug Discov*, 1(8):621–636.

Index

ACE□□□, 181
ACHILLES, 214
adaboost, 177
agnostic SQL, *see* SqlRender
analysis implementation, 69
APHRODITE, 105
arachne, 260
ATHENA, 35
ATLAS, 70, 182
    characterization features, 133
    □□□□□□, 72
    □□□□□□□□□, 71
    □□□□□□, 71
    □□□□□□□□, 71
    □□□, 71
    □□□□□□, 72
    □□□□□□, 70
    □□□□□□, 72
    □□□□□□□, 71
    □□□□□□, 71
    □□□□□□□□, 71
    □□□□□□□, 71
    □□□□□, 71
    □□□, 71
    □□, 71
    □□□□□□□, 71
attrition diagram, 168
AUC, 180

back-propagation, 178
balance, *see* covariate balance
baseline time, 123
best practice for network research, 261
between-database heterogeneity, 243
bigknn, 177
Bill of Mortality, 33

caliper, 147
    scale, 156
case-control design, 147
case-crossover design, 148
case-time-control design, 148
CDM , *see* Common Data Model
characterization, 63, 123
    cohort, 124
    database level, 123
    treatment pathways, 124
classification concept, 37
Clem McDonald, 212
clinical equipoise, 147
cohort
    entry event, 104
    exit criteria, 104
    inclusion criteria, 104
    probabilistic design, 105
    rule-based design, 104
cohort method, 145
colliders, 147
Common Data Model, 19
    □□□□□□□□, 21
    □□□□□□, 21
    □□□□□□□, 19
    □□□□□□□, 19
    □□□□□, 19
    □□□□□□□, 21
    □□□□, 21
    □□□□□, 21
    □□□□□□, 21

□□□□□, 19
community, 5, 212
comparative effect estimation, 145
comparative effectiveness, *see* comparative effect estimation
comparator cohort, 145
concept, 35
  class, 36
  code, 38
  hierarchy, 41
  identifier, 35
  mapping, 39
  relationship, 39
concept set, 105
conditioned model, 156
confidence interval calibration, 242
confounder, 146
control hypotheses, 72
convolutional neural network, 178
counterfactual, 145
covariate balance, 147
  example, 168
Cox proportional hazards model, *see* Cox regression
Cox regression, 146
cross-validation, 176
Cyclops, 175

data profiling, *see* White Rabbit
data quality, 213
DatabaseConnector, 81
  creating a connection, 89
  querying, 90
decision boundary, 175
decision tree, 177
deep learning, 178
descriptive statistics, *see* characterization
design considerations for network research, 258
direct effect estimation, 145
disease natural history, *see* characterization
domain
  concept, 36
drug utilization, 124

ETL, *see* □□□□□□□□□□□ETL□
  implementations, 57
  quality control, 57
  □□□□□, 215
ETL design, *see* Rabbit-In-A-Hat
evidence quality, 211

FAIR, 15
feature analyses, 128
FeatureExtraction, 134
forum, 7

gradient boosting, 176

high correlation, 160
hyper-parameter, 176

incidence, 124
  proportion, 125
  rate, 125
index date, 123
instrumental variables, 147

k-nearest neighbors, 177
Kaplan-Meier plot, 170

LASSO, 175
logistic regression, 146, 175
logistical considerations for an OHDSI network study, 258

methods library, 72
minimum detectable relative risk (MDRR), 168
mission, 4
model viewer app, 198

naive bayes, 177
nesting cohort
  case-control design, 147
network study, 257
neural network, 178
no free lunch, 175

objectives, 5
OHDSI Methods Benchmark, 249
OHDSI SQL, *see* SqlRender

outcome cohort
    case-control design, 147
    case-crossover design, 148
    cohort method, 145
    SCCS design, 149
    self-controlled cohort design, 147

p-value calibration, 242
Pallas system, 33
patient-level prediction, 65
perceptron, 178
person-time, 125
phenotype library, 106
PheValuator, 227
Poisson regression, 146
population-level estimation, 64, 145
post-index time, 123
power, 168
preference score, 147
    example, 167
propensity model, 146
    example, 167
propensity score, 146
    matching, 146
    stratification, 146
    trimming, 156
    weighting, 146
python, 176, 177

quality improvement, *see* characterization
Query Library, 81
QueryLibrary, 94

R, 81
    installation, 74
Rabbit-In-A-Hat, 48
random forest, 176
randomized trial, 146
recurrent neural networks, 178
regularization, 175
reliable evidence, 212
ROC, 180
running network research, 259

safety surveillance, 145
self-controlled case series (SCCS) design, 149
self-controlled cohort design, 147
sensitivity analysis, 243
sklearn, 176
source code mapping, *see* Usagi
SQL, 81
SQL Query Library, *see* Query Library
SqlRender, 81
    debugging, 88
    parameterization, 82
    supported functions, 84
    translation, 83
standard concept, 37
Standard SQL Dialect, *see* SqlRender
standardized vocabularies, 33
    download, 35
    search, 35
stratified model, conditioned model□□□158
strongly ignorable, 146
structured query language, *see* SQL
study diagnostics, 255
study feasibility
    single study, 255
study-a-thon, 14
supervised learning(□□□□), 174
survival plot, *see* Kaplan-Meier plot
system requirements, 74

target cohort
    case-control design, 147
    case-crossover design, 148
    cohort method, 145
    SCCS design, 149
    self-controlled cohort design, 147
tools deployment, 77
    Amazon AWS, 78
    Broadsea, 78
treatment utilization, *see* characterization
TRIPOD, 173

Usagi, 53

variable ratio matching, 146

variance, 176
vignette, 74
vision, 4
vocabulary, 36

White Rabbit, 45
workgroups, 9

xgboost, 176

□□□□□□□□□, 173
□□□□□□□□□□, 174
□□□□□□□, 181
□□□□□□□, 180
□□□□□□□□□, 13
    open standards, 14
    □□□□□□□, 14
    □□□□□□□□□□, 14
    □□□□□□□, 14
□□□□□□□□□, 180
□□□, 174
□□□□, 103
□□□□□□, 103
□□□□□□
    □□□□□□□□□, 9
□□□□□
    □□, 42
□□□□□□, 103
□□□□□□, *see* □□□□
□□□□□□□□□, 226
□□□□□□□□□, 173
□□□□□
    □□□□□□□□□, 217
    □□□□, 214
    □□□□□□□, 214
    □□□, 214
    □□□, 214
    □□, 214
    □□, 214
□□□□□□□□, *see* □□□□
□□□□□□□□□□□, 240
□□□□□□□□□, 179
□□□□□, 253
□□□□□□□□□□□, 241
□□□, 174
□□□□□, 173
□□□□□□□□□□□□□, *see* Common Data Model

□□□□□□□, 173
□□□□□, 173
□□□□□□□□, 178
□□□□, 179

□□□, 179
□□□, 179
□□□, 174
□□□□□□□□
    □□□□□□□, 24
    □□, 21
□□□, 174
□□□□□□□□□□□□, 242
□□□□□, 241

□□□□□□□, 173
□□, 179
□□□□□□, 239
□□
    □□□□, 178
    □□□□, 178
    □□□□□, 179
    □□□□□, 179
□□□□, 173
□□□□□, 66, 175

□□□, 179
□□□□, 45
□□□, 179
□□□, 179
□□□□□□□□, 257
□□□□□□□, 254
□□□□, 239
□□, 179

□□□□□□, 173
□□□, 103
□□□□□□□, 66
□□□□□□□, 173
□□□, 180

□□□-□□□□□□□□□, 180
□□□□□, 179