– – OHDSI

Observational Health Data Sciences and Informatics (OHDSI)

2025-03-03

Contents

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

I OHDSI □□□□□□ 1

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

2 –□□□□□– □□□□□□□□□□ 7
2.1 □□□□□□□□ 7
2.2 □□□□□□□□□□□ 10
2.3 □□□ 11

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

□4□ –□□□□□– □□□□□□□□□ 17

4 □□□□□□□□ 19
4.1 □□□□ 19
4.2 □□□□□□□□□ 21
4.3 CDM□□□□□□□ 23
4.4 □□□□ 28
4.5 □□□ 28
4.6 □□□□ 28

5 □5□ –□□□□□– □□□□□□□□□ 31
5.1 □□□□□□□□□□□□□□□□□□□□□□□ 31
5.2 □□□□□ 33
5.3 □□ 37
5.4 □□ 39
5.5 □□□□□□□□ 40
5.6 □□□□□ 40
5.7 □□□ 41
5.8 □□ 41

6 □6□ –□□□□□– □□□□□□□□□ 43
6.1 □□□□ 43
6.2 □□□□2: □□□□□□□□□□□ 49
6.3 □□□□ 3: ETL□□□ 51
6.4 □□□□ 4: □□□□ 51
6.5 ETL□□□□THEMIS 51
6.6 CDM□□□ETL□□□□□□□ 52
6.7 ETL□□□□□□□□□□ 52
6.8 □□□ 52
6.9 □□ 52

□7□ –□□□□□– □□□□□ 55

7 □□□□□□□□□ 57
7.1 □□□□ 57
7.2 □□□□□□□□ 58
7.3 □□□□□□□□ 59
7.4 □□□□□□□□□□ 59
7.5 □□□□□□□ 60
7.6 □□□ 60
7.7 □□ 60

8 □8□ –□□□□□– OHDSI □□□□□ 63
8.1 □□□□□ 63
8.2 □□□□ 64
8.3 ATLAS 64
8.4 □□□□□□□□□ 66

8.5 □□□□ 68
8.6 □□□ 69

9 □9□ –□□□□□– SQL□R 71
9.1 SqlRender 72
9.2 DatabaseConnector 78
9.3 CDM□□□□□ 80
9.4 □□□□□□□□□□□□□□□□ 83
9.5 QueryLibrary 84
9.6 □□□□□□□□□□□ 85
9.7 SQL□R□□□□□□□□□□□ 85
9.8 □□□ 90
9.9 □□□□ 90

10 □10□ –□□□□□– □□□□□□□ 93
10.1 □□□□□□□ 93
10.2 □□□□□□□□□□□□□ 94
10.3 □□□□□□□□ 95
10.4 □□□□□□□□□ 95
10.5 □□□□□□□□□□ 95
10.6 □□□□□□□□□□ 96
10.7 ATLAS□□□□□□□□□□□ 97
10.8 SQL□□□□□□□□□□□□ 101
10.9 □□□□ 109
10.10□□ 109

11 □11□ –□□□□□– □□□□ 111
11.1 □□□□□□□□□□□□□□ 111
11.2 □□□□□□□□ 111
11.3 □□□□ 112
11.4 □□□ 112
11.5 □□□□□□□□□□ 113
11.6 ATLAS□□□□□□□□□□□□□□□ 113
11.7 ATLAS□□□□□□□□□□□□□ 115
11.8 R□□□□□□□□□□□□ 120
11.9 ATLAS□□□□□□□□□□□□□□□ 124
11.10ATLAS □□□□□□□□□□ 125
11.11□□□ 131
11.12□□ 131

12 □12□ –□□□□□– □□□□□□□□ 133
12.1 □□□□□□□□□□ 133
12.2 □□□□□□□□□□□□ 135
12.3 □□□□□□□□ 136
12.4 □□□□□□□□ 136

12.5 □□□□□□□□□□□□□□ 137
12.6 □□□□□□□□□□□ 137
12.7 ATLAS□□□□□□□□□□□ 139
12.8 R□□□□□□□□□□□ 141
12.9 □□□ 147
12.10□□ 147

13 □13□ –□□□□□– □□□□□□□ 151
13.1 □□□□ 151
13.2 □□□□□ 152
13.3 □□□□□□ 153
13.4 □□□□□□□□ 156
13.5 □□□□□□□□□□□□□ 158
13.6 ATLAS□□□□□□□ 160
13.7 R□□□□□□ 164
13.8 □□□□□□□□□□□□□ 176
13.9 □□□ 176
13.10□□ 176

II –□□□□□– □□□□ 179

14 □□□□□□□ 181
14.1 □□□□□□□□□□□□□□ 181
14.2 □□□□□□□ 182
14.3 □□□□□□□□□□□ 183

15 –□□□□□– □□□□□ 185
15.1 □□□□□□□□□□□ 185
15.2 □□□□□□□□□□ 188
15.3 □□□□□□ACHILLES 191
15.4 □□□□□□□□□□□□□ 192
15.5 □□□ 194
15.6 □□ 194

16 –□□□□□– □□□□□□ 197
16.1 □□□□□□□□□□□□ 197
16.2 □□□□□□□□□□□□ 197
16.3 □□□□□□□□□□ 198
16.4 PheValuator 199
16.5 □□□□□□□□□□ 203
16.6 □□ 203

17 –□□□□□– □□□□□□□□□□□ 205
17.1 □□□□□□□□□□ 205

17.2 □□□□□□□□□□□□□□□□□□□□□□□□ 206
17.3 □□□□□□□□□□□□□□ 207
17.4 □□□ 208

18 –□□□□□– □□□□□□ 209
18.1 □□□□□□□□□□ 209
18.2 □□□□□□□□□ 209
18.3 □□□□□□□□□□□□□ 212
18.4 OHDSI□□□□□□□□□□□□□□ 214
18.5 □□□ 215

III OHDSI□□ 217

19 –□□□□□– □□□□□ 219

20 –□□□□□– OHDSI □□□□□□□□□ 223
20.1 OHDSI □□□□□□□□□□□□□ 223
20.2 OHDSI□□□□□□□□□ 223
20.3 OHDSI□□□□□□□□□□□□□□□□□□□□□ 226
20.4 □□□ 227

Appendix 227

A –□□□□□– □□□ 229

B □□□□□□ 233
B.1 ACE□□□ 233
B.2 ACE□□□□□□□□□□□□□□□ 234
B.3 □□□□□□□□AMI□ 237
B.4 □□□□□ 238
B.5 □□□□□□□□□□□□□□□□□□□□□□□□□□□ 239
B.6 □□□□□□□□□□□□□□□□□□□□□□□ 241
B.7 □□□□□3□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□ 244
B.8 ACE□□□□□□□ 245
B.9 □□□□□□□□□□□□□□□□ARB□□□□ 245
B.10 □□□□□□□□□□□□□□□□□□□□□□□□□□ 247
B.11 □□□□□□□□□□□□□□□□□□□□□□□□□□□DCCB□□□□ 247
B.12 □□□□□□□□□□□□□□□□□□□□□□□□□□□□NDCCB□□□□ 248
B.13 β□□□□□ 248
B.14 □□□□□□□□□ 249
B.15 □□□□□□□□□□□□□□ 249
B.16 α1□□□□□ 249

C □□□□ 251
C.1 ACE□□□□□□□□□□□□□□□□□□□□□□□□□□ 251

D □□□□□□□□□□□ **255**

E □□□ **257**
E.1 □□□□□□□□ 257
E.2 □□□□□□□□□□□□ 260
E.3 ETL (Extract-Transform-Load) 261
E.4 □□□□□□□□□□□□ 262
E.5 SQL□R 262
E.6 □□□□□□□ 264
E.7 □□□□□ 269
E.8 □□□□□□□□ 276
E.9 □□□□□□□□ 280
E.10 □□□□□ 282
E.11 283

Bibliography **285**

Index **293**

□□□□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

□□□□□□□□□□□□□□□□ & □□□□□□□□□□□□

□□□□□□□□□□□□□□□□

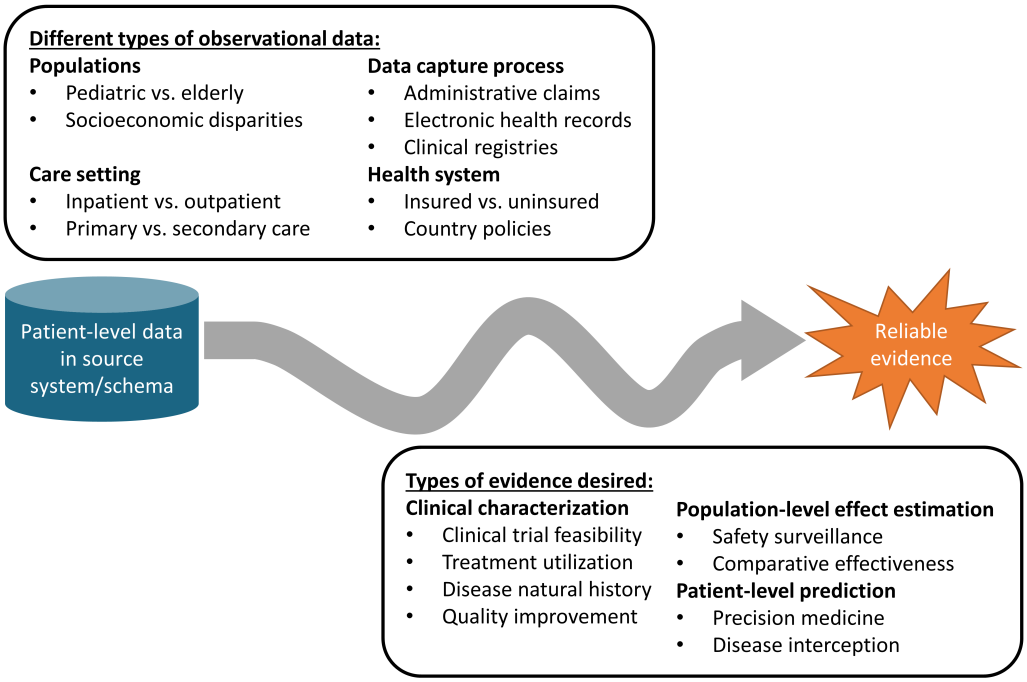


Figure 1.1: □□□□□□□□□□□□□□□

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

1.2.2

□□□

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

Chapter 2

— —

□□□□□□□□□□□□□□□□ & □□□□□□□□□□□□

□□□□□□□□□□□□ - □□

OHDSI□□□□□□□□□□□□□□□□

□□□□□□□□□

□□□□□□□□	□□	□□□□□
Natural Language Processing	OHDSI□□□□□□□

2.1.5 OHDSI

OHDSI □□□□□□□□□□

2.3

- OHDSI□□□

Chapter 3

— —

□□□□□: □□□□□□□□□□□□

OHDSI□□□□□□□□□□□

3.3

OHDSI□□□OMOP□□□□□□□□□□□□
4□□□□□□□□□□□□□□□□□□□□□□□□ 5□□□□□□□□□□□□□□□□□□□

3.7.4

□□□□□□□□OMOP□□□□□□□□OHDSI□□□□□□□□□□□□□□□□

4 – –

Chapter 4

□□□□: *Clair Blacketer*

□□□□

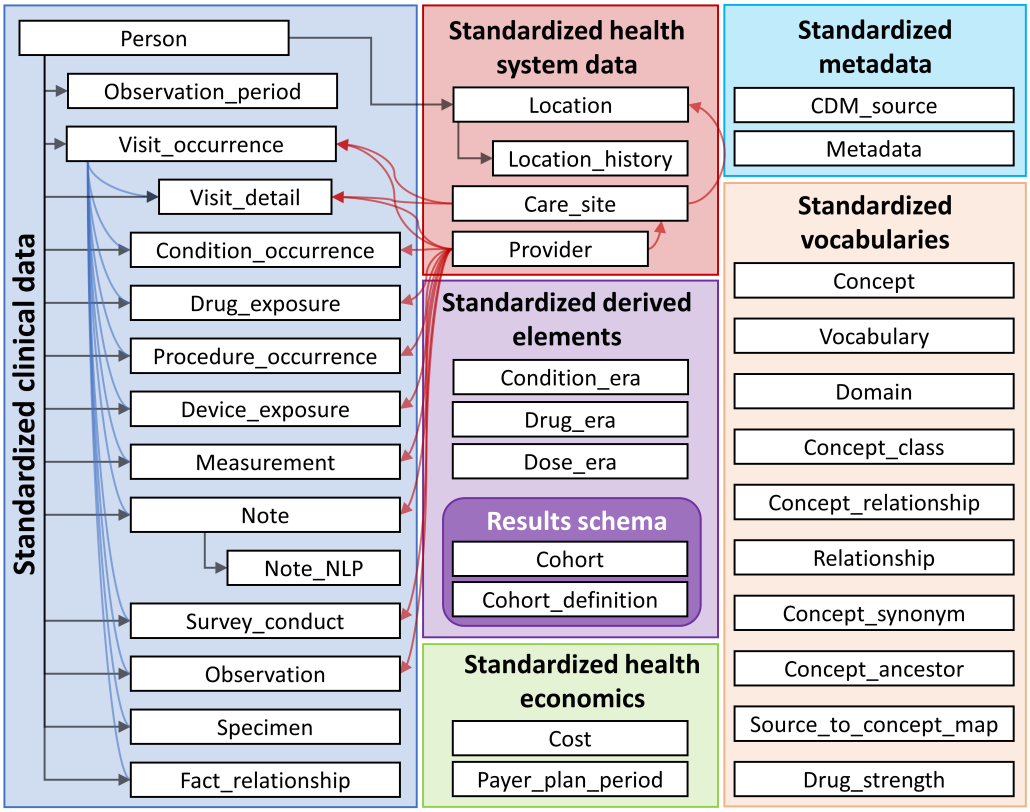


Figure 4.1: CDM□□□□□6.0□□□□□□□□

- □□□□□□□□: □□□□□□□□□□□

□□□□□□	□□□□ID	□□□□□□	□□□□ID
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□□□

4.3.1 :

□□□□□□□

Table 4.3: PERSON□□□□□

□□	□	□□
PERSON_ID	1	PERSON_ID□□□□□□□□

□□□□□□□ID	□□□	□□□	□□□
80	2011-01-06	2011-01-06	□□□□
90	2012-01-06	2012-01-06	□□□□
100	2013-01-07	2013-01-07	□□□□
101	2013-01-14	2013-01-14	□□□□
102	2013-01-17	2013-01-24	□□□□

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

Table 4.5: OBSERVATION_PERIOD□□□□□

□□	□	□□
OBSERVATION_ PERIOD_ID	1	□□□□□□□□□□□□□□□□□□□□

□□	□	□□
VISIT_TYPE_ CONCEPT_ID	32034	□□□□□□□□□□□□□□□□□□□□□□□□□□□□EHR□
PROVIDER_ID	NULL	□□□□□□□□□□□□□□□□□□□□

□□	□	□□
CONDITION_START_DATE	2010-01-06	□□□□□□□□□□□□□□
CONDITION_START_DATETIME	2010-01-06 00:00:00	□□□□□□□□□□□□□□

□□ **4.2.** □□□□2015□1□1□□□□□□□□□□□□□□□□□□□□□□□□□□□□2019□7□1□□□□□□□□□

□□ **4.3.** □□□□2019□5□1□□□□□□□□□□200 MG□□□□□□NDC□□□□76168009520□□30□

□□□□

□□□3□□□□□□□□□□□□□□@ref□installR□□□□□□□□□□□□□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□□□□□□□□□□□□

```
library(DatabaseConnector)
connection <- connect(connectionDetails)
sql <- "SELECT *
FROM @cdm.condition_occurrence
LIMIT 1;"
result <- renderTranslateQuerySql(connection, sql, cdm = "main")
```

□□ **4.4.** SQL□R□□□□□□□□□□□□□□□□□□□□□□□ID192671□□□□□□□□□□□□□□□□□□□□□□

□□ **4.5.** SQL□R□□□□□□□□□□□□□□□□□□□

Chapter 5

5 — —

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

OMOP□□□OHDSI□□□□□□□

□□□□□□□□□□□□□□□□□□□

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□□□□□□□□□□□□□□□□□□□□□□□□ATHEN

OMOP CDM□□zip□□□□□

5.1.3 :

OHDSI□□□□□□□□□□□□□□□□□□□

5.2.3

□□□□□□□□DOMAIN_ID□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□CONCEPT_ID□□□
□□□□□□□□□□□□Pallas□□□OM

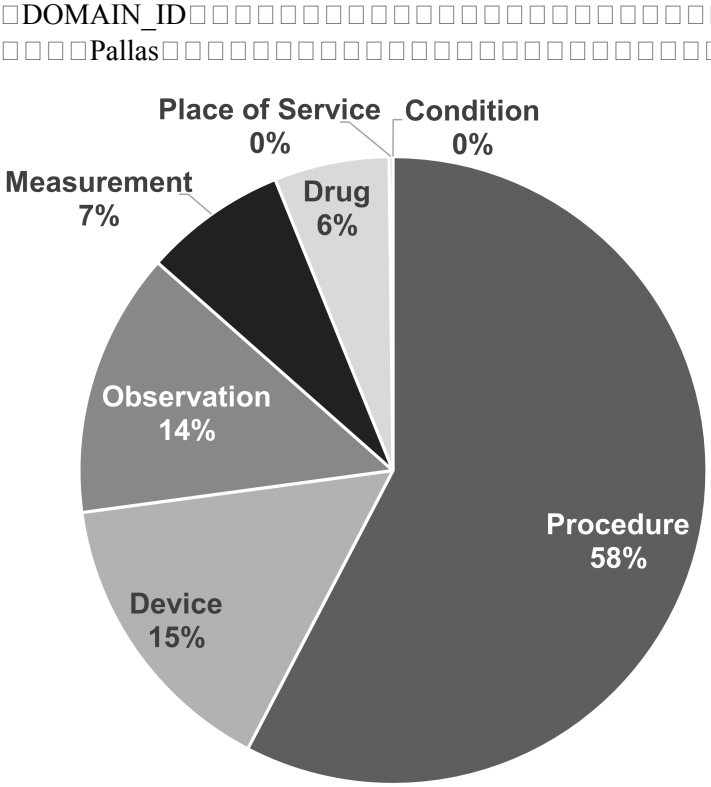


Figure 5.3: □□□□□□□□□□CPT4□□□HCPCS□□□□□□□□□□□□□□□□□□□

| □□□□□□

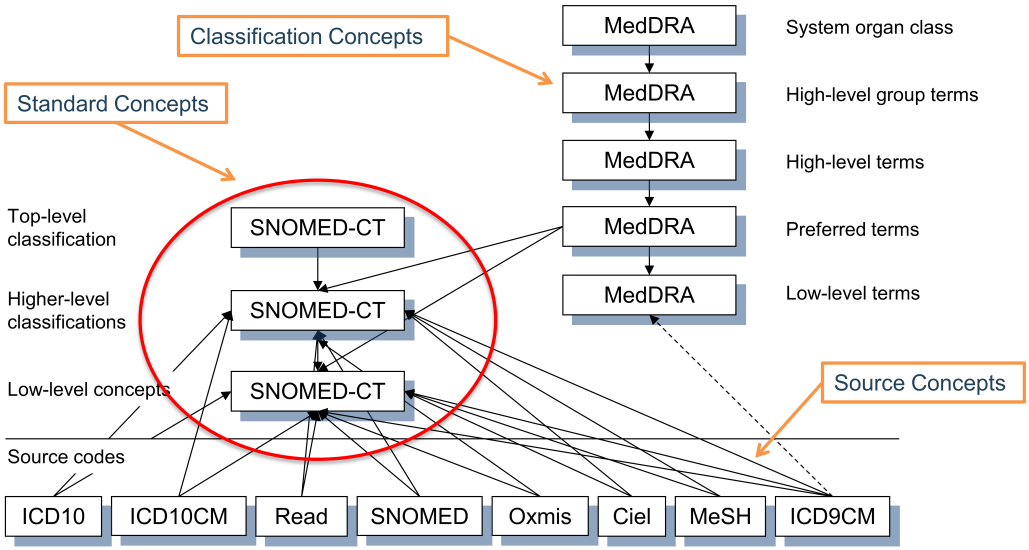

Classification Concepts
Standard Concepts
MedDRA
MedDRA
MedDRA
MedDRA
MedDRA
System organ class
High-level group terms
High-level terms
Preferred terms
Low-level terms
Top-level classification
Higher-level classifications
Low-level concepts
SNOMED-CT
SNOMED-CT
SNOMED-CT
Source Concepts
Source codes
ICD10
ICD10CM
Read
SNOMED
Oxmis
Ciel
MeSH
ICD9CM

Figure 5.4: □□□□□□□□□□

- VALID_START_DATE: □□1
1-1□ - VALID_END_DATE: □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
- INVALID_REASON: “U” - □□□□□□□□□□□□□□□□□□□□□□□□ - □□:
□□□□□□□□□□□□□□□□□□□

□□□□□□□□□

Table 5.5: □□□□Finding site of□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

CONCEPT_ID_1	CONCEPT_ID_2
4000504 “Urethra part”	36713433 “□□□□□□□”
4000504 “Urethra part”	433583 “□□□□□□”
4000504 “Urethra part”	443533 “□□□□□□□□”
4000504 “Urethra part”	4005956 “□□□□□□□□”

□□□□□□□□□□□□□□□□□□□

5.5

DOMAIN_ID□VOCABULARY_ID□CONCEPT_CLASS_ID□□□□CONCEPT□□□□□□□□□□CON
VOCABULARY□□VOCABULARY_REFI

5.6

5.6.1

OMOP CDM□□□□□□□□□□□□□□□□□□□

5.7

- □□□□□□□□□□□□□□□□□OMOP□□□□□□□□□□□□□□□□

Chapter 6

6 — —

□□□□: *Clair Blacketer & Erica Voss*

6.1

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

ETL□□4□□□□□□□□□□□□□□□

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

□□□OHDSI□□□□□□□□□□□□□□□□□□□□□
□□□□ 1: ETL□□□□□

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: □□□□□□□□□□□□□□□□□□□□

	A	B
1	Sex	Frequency
2	2	61491
3	1	35401
4	List truncated...	

Figure 6.4: □□□□□□□□□□□

□□□□

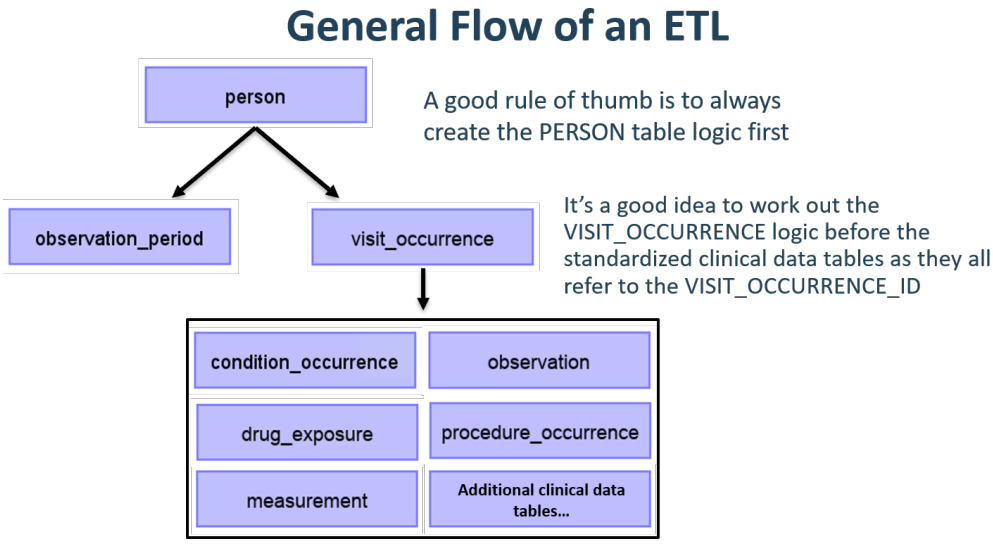


Figure 6.5: ETL□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

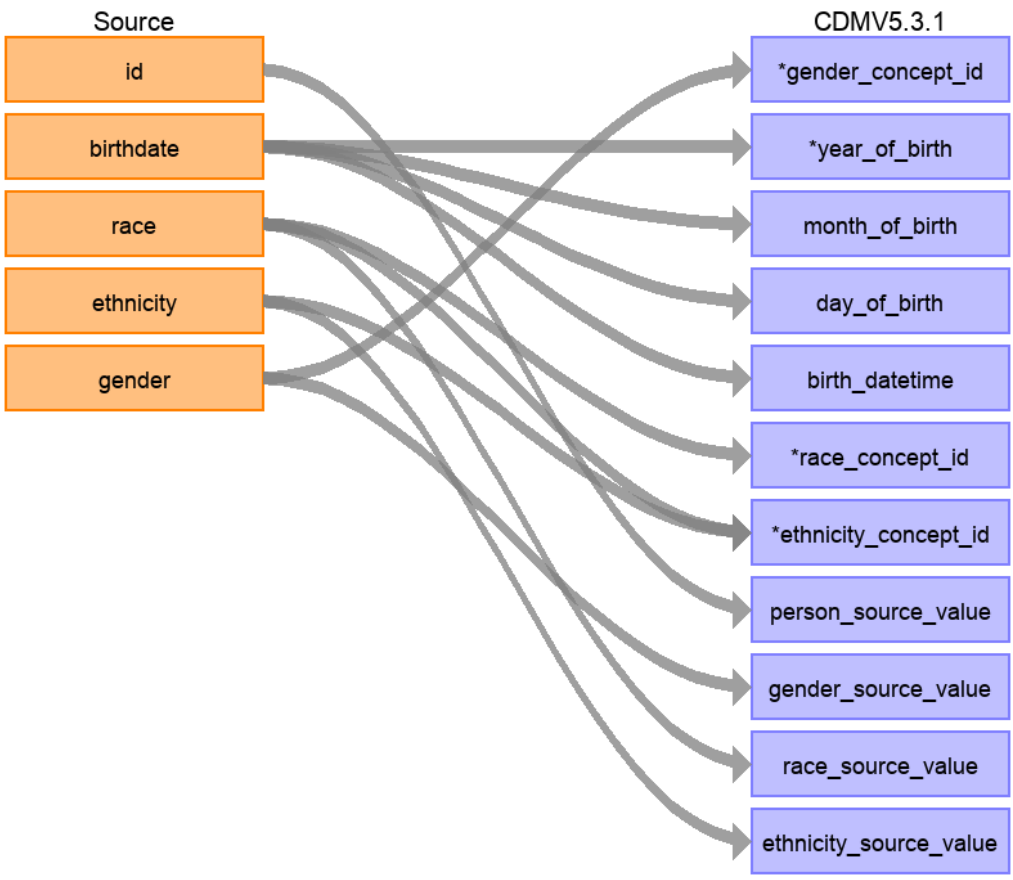


Figure 6.6: Synthea Patients□□□□□□□CDM PERSON□□□□□□□□□□□□□□

□□(□tab:syntheaEtlPerson) Synthea Patients□□□□□CDM PERSON□□□□□□□□□□□□□ETL

□□□□□□□	□□□□□□□□□□□□□□□□□
PERSON_ID	□□□□□ PERSON_ID□□□□□□□□□□□□□□□□□□□□□□□id□□var □□□□id□□□□□□□□□□□□□□□□□□□□□□□□□□□□
GENDER_CONCEPT_ID gender	□□□□M□□□□□GENDER_CONCEPT_ID□8507□□
YEAR_OF_BIRTH birthdate	□□□□□□

6.2 2:

OMOP□□□CDM□□□□ 10CM□□□□□□□□□□□□□□□□□□SNOMED□□□□□□□□□□□□□□□□□□□□□□□□□□□relationship_id =□Maps to□□□□CONCEPT_RELATIONSHIP□□□□□□□□□□□□□□□□□□□□□□□□□□□□ICD-10CM□□□□I21□□□□□□□□□□□□□□□□□□□

□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

1. □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
2. Usagi□□□□□□□□□□□□□□□□□□□

6.3 3: ETL

□□□□□□□□□□□□□□□□□□□□□□ETL□□□□□□□□□□□□□□□□□□□□□□□□□□□□ETL□□□
6.6 □□□□□□□□□□□□

□□ETL□□□□□□□□□□□□□□□□□

□□□ETL□□□□□□□□□□□□□□□□□□□□□□□□□

- ETL-Synthea - Synthea□□□□□□□□□□□□□□□□□□□□□SQL□□□□
  - https://github.com/OHDSI/etl-synthea
- ETL-CDMBuilder - □□□□□□□□□□□□□□□□□□□□□□□□□□□□.NET□□□□□□□□□
  - https://github.com/OHDSI/etl-cdmbuilder
- ETL-LambdaBuilder - AWS lambda□□□□□□□□□□□□
  - https://github.com/OHDSI/etl-lambdabuilder

□□□□□□□□□□□□□□□□□□□□□□□□□□□□”□□□”ETL□□□□□□□□□□□□□□□□□□□

□□□□□□□□□

C) □□□□□□□□□□□□□□□□□□□□□□□□
D) □□□□□□□□□□□□□□□□

□□ **6.2.** □□□□OHDSI□□□□□□□□□□□□6.2□□□PERSON□□□□□□□□4□□□□□□□□□□

□: PERSON□□□□□

□	□
PERSON_ID	A123B456
GENDER_CONCEPT_ID	8532
YEAR_OF_BIRTH	NULL
MONTH_OF_BIRTH	NULL
DAY_OF_BIRTH	NULL
RACE_CONCEPT_ID	0
ETHNICITY_CONCEPT_ID	8527
PERSON_SOURCE_VALUE	A123B456
GENDER_SOURCE_VALUE	F
RACE_SOURCE_VALUE	WHITE
ETHNICITY_SOURCE_VALUE	□□□□□□□□□

□□ **6.3.** VISIT_OCCURRENCE□□□□□□□□□□□□□□□□□□□□□Synthea□□□□□□□□□□□□□
PATIENT□START□END□□□□□□□□□□□□□□□□□□PERSON_ID□□□□□□□□□END□□□□□

- MIN(START)□VISIT_START_DATE□□□□□
- MAX(END)□VISIT_END_DATE□□□□□
- “IP”□PLACE_OF_SERVICE_SOURCE_VALUE□□□□□

□□□□□□□□□6.8□□□□□□□□□□□□□□□□□□□□□□CDM□□□□□□VISIT_OCCURRENCE□

Data Output Explain Messages Notifications Query History

	id character varying (1000)	start date	stop date	patient character varying (1000)	encounterclass character varying (1000)
1	12	2004-09-26	2004-09-27	11	inpatient
2	13	2004-09-27	2004-09-30	11	inpatient

Figure 6.8: □□□□□□□□□

□□□□□□□□□□E.3□□□□□□□□□□□□□□

7 – –

Chapter 7

□□□□□□□□: *David Madigan*

OHDSI□□□□□□□□

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

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

7.3

□□□□□□□□

7.4.2

□□□□□□□□□□□□□□□Chapter 12□□□□□□□□□□□□□□□

2. □□□□□□□□□□□□□□□□□□□□□□□□□GI□□□□□□□□□□□□□□□□□□
3. □□□□□□□□□□□□□□□□□□□□□□□□□GI□□□□□□□□□□□□□□□□□

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

□□□□□□□□□□Appendix E.4□□□□□□□

Chapter 8

8 – – OHDSI

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

OHDSI□□□□□

Library□□□□□

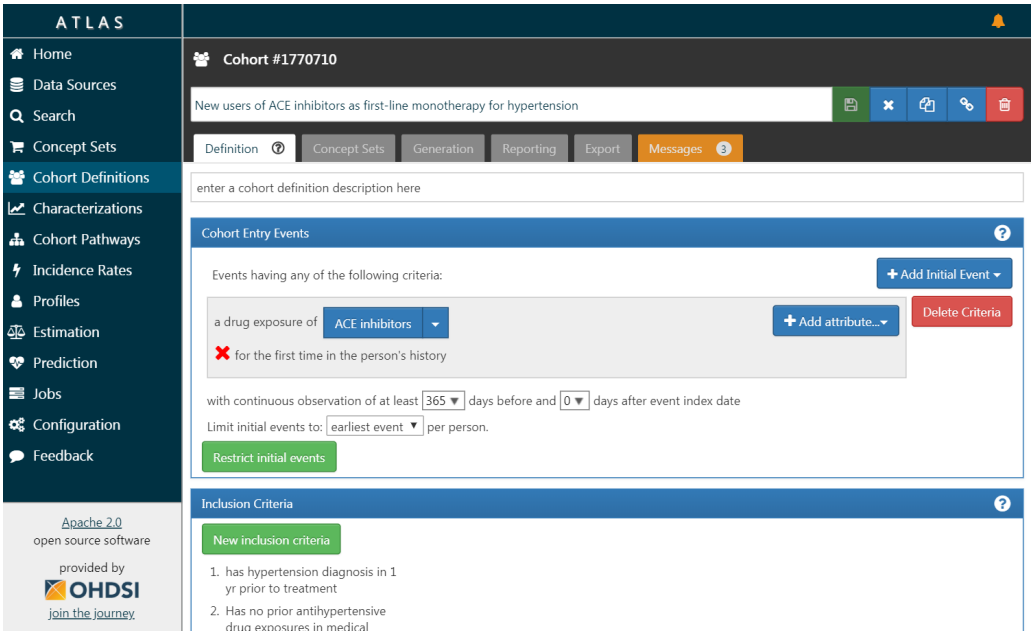


Figure 8.3: ATLAS□□□□□□□□□□□

□□□□□□□□ □□□□□□□□□□□□

8.3.2

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

Figure 8.4: OHDSI□□□□□□□□□□□□□□□□□□□□

8.4.3

R□□□□□

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

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
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.avc9409echildren ;"
## attr(,"sqlDialect")
## [1] "oracle"
```

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

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

□□□□□□□□□□□□

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

□□□□□□□SQL Server□□□□□□□□□□□□□□□□□□□

9.2.1

□□□

```
querySql(conn, "SELECT TOP 3 * FROM person")
```

```
##   person_id gender_concept_id year_of_birth
## 1         1               8507          1975
## 2         2               8507          1976
## 3         3               8507          1977
```

```
executeSql(conn, "TRUNCATE TABLE foo; DROP TABLE foo;")
```

□□□□□

```
SELECT COUNT(*) AS person_count FROM @cdm.person;
```

PERSON_COUNT
26299001

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

```
SELECT AVG(DATEDIFF(DAY,
                    observation_period_start_date,
                    observation_period_end_date) / 365.25) AS num_years
FROM @cdm.observation_period;
```

NUM_YEARS
1.980803

□□□□□□□□□□□□□□□□□□□

```
        INNER JOIN @cdm.observation_period
            ON person.person_id = observation_period.person_id
        ) age_computed
    )
SELECT MIN(age) AS min_age,
    MIN(CASE
            WHEN order_nr < .25 * n
                THEN 9999
            ELSE age
            END) AS q25_age,
    MIN(CASE
            WHEN order_nr < .50 * n
                THEN 9999
            ELSE age
            END) AS median_age,
    MIN(CASE
            WHEN order_nr < .75 * n
                THEN 9999
            ELSE age
            END) AS q75_age,
    MAX(age) AS max_age
FROM ages
CROSS JOIN (
    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
```

□□□□□□□□□□□□□□□□□□□

```
SELECT TOP 10 condition_source_value,
  COUNT(*) AS code_count
FROM @cdm.condition_occurrence
GROUP BY condition_source_value
ORDER BY -COUNT(*);
```

CONDITION_SOURCE_VALUE	CODE_COUNT
4019	49094668
25000	36149139
78099	28908399
319	25798284
31401	22547122
317	22453999
311	19626574
496	19570098
I10	19453451
3180	18973883

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

9.4

□□□CDM□□□□□□□□S

```
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

□□□□□□□□□□□□□□□□□□□

```
  ON drug_concept_id = descendant_concept_id
INNER JOIN @cdm.concept ingredient
  ON ancestor_concept_id = ingredient.concept_id
WHERE LOWER(ingredient.concept_name) = 'ibuprofen'
  AND ingredient.concept_class_id = 'Ingredient'
  AND ingredient.standard_concept = 'S';
```

PRESCRIPTION_COUNT
26871214

9.5 QueryLibrary

QueryLibrary□□CDM□□□□□□SQL□□□□□□

9.6

9.6.1

□□□□□□ACE□□□□□ACEi□□□□□□□□□□□□□□□□Slater et al. (1988) □□□□□□ACEi□□□□□

□□□□□□□□□□□□□□□□□1□□□□□□□□□□□□□□□□□

9.7.1

□□□□□□□□□□□□□□□COHORT□□□□□□□□□□□□□

```
sql <- "
INSERT INTO @cohort_db_schema.@cohort_table (
  cohort_definition_id,
  cohort_start_date,
  cohort_end_date,
  subject_id
)
SELECT 1 AS cohort_definition_id,
  cohort_start_date,
  cohort_end_date,
  subject_id
FROM (
  SELECT drug_era_start_date AS cohort_start_date,
    drug_era_end_date AS cohort_end_date,
    person_id AS subject_id
  FROM (
    SELECT drug_era_start_date,
      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_EXPOSURE

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□□□□□□□□□□□□□

9.7.3

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

```
sql <- "
WITH tar AS (
  SELECT concept_name AS gender,
    FLOOR((YEAR(cohort_start_date) -
          year_of_birth) / 10) AS age,
```

```
    subject_id,
    cohort_start_date,
    CASE WHEN DATEADD(DAY, 7, cohort_start_date) >
      observation_period_end_date
    THEN observation_period_end_date
    ELSE DATEADD(DAY, 7, cohort_start_date)
    END AS cohort_end_date
  FROM @cohort_db_schema.@cohort_table
  INNER JOIN @cdm_db_schema.observation_period
    ON subject_id = observation_period.person_id
      AND observation_period_start_date < cohort_start_date
      AND observation_period_end_date > cohort_start_date
  INNER JOIN @cdm_db_schema.person
    ON subject_id = person.person_id
  INNER JOIN @cdm_db_schema.concept
    ON gender_concept_id = concept_id
  WHERE cohort_definition_id = 1 --
)
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_

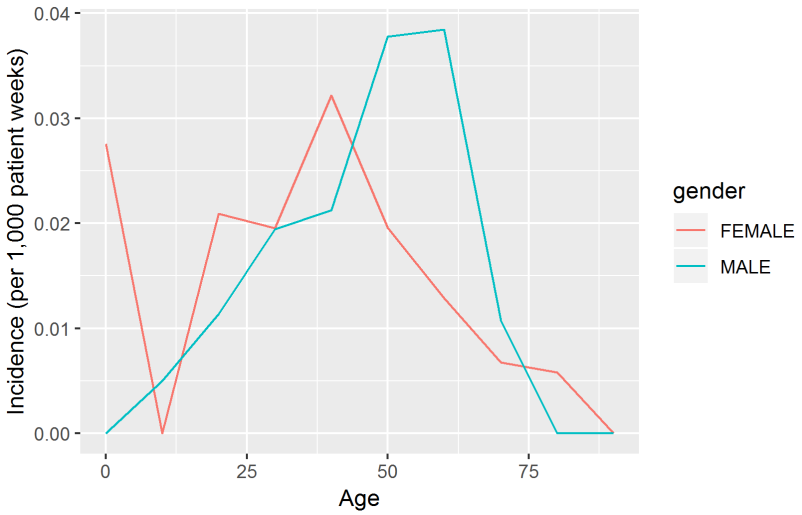
SQL□□□□□□□□□□□□□□□□□□□□□□□□□□□□SQL□□□□□□□□□□□□□□□□□□□□□R□□□□□

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

□□□□□□□□□□□□□□□□□□□

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

disconnect(conn)
```

9.7.5

OHDSI SQL□DatabaseConnector□□□SqlRender□□□□□□□□□□□□□□□□□□□□□□□□□□□□□OHD

□□□□□□□□□□□□□□□□□□□□□□□□□□SQL□□□□□□□□□□□□□□□□□□□□□□□□□□□□□ATLAS□
SQL□□□□□□SqlRender□□□DatabaseConnector□□□□□□□□□□□□□□□□

9.8

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

□□ **9.2.** SQL □□□ R □□□□□□□□□□□□ 1 □□□□□□□□□□□□

Chapter 10

10 – –

□□□□□: *Kristin Kostka!*

□□□□□

10.2

□□□□□□□

10.3

□□□□□□□□□

□□□□□□□

10.7 ATLAS

□□ATLAS□□□□□□□□ Cohort Definitions □□□□□□□□□□□□□

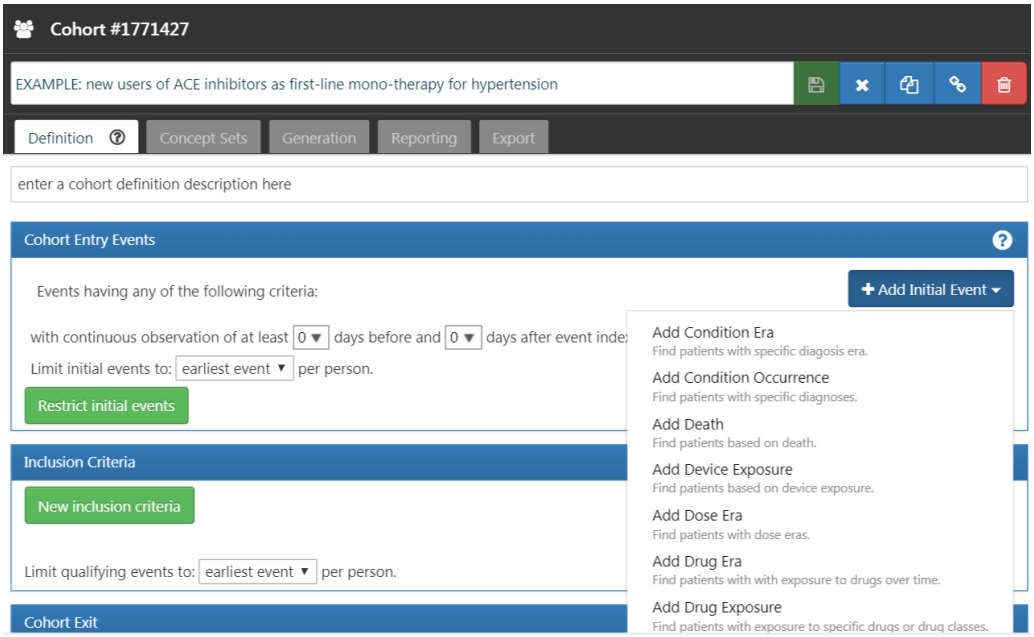


Figure 10.5: □□□□□□□□□

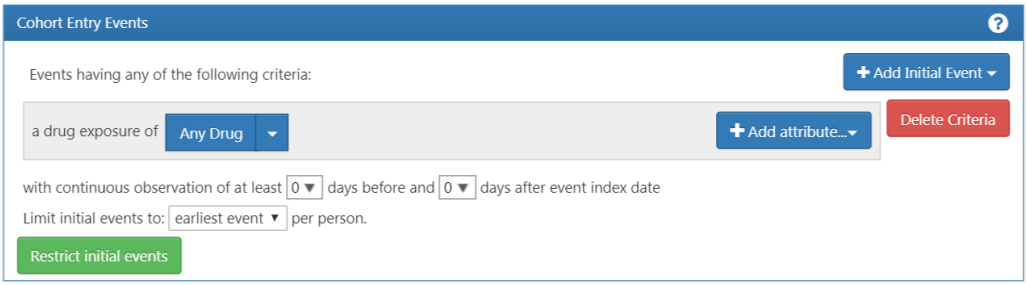


Figure 10.6: □□□□□□□

□□□□1: □□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□

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□□□□□□□□□□□□□□□□□□□□□□□□□□

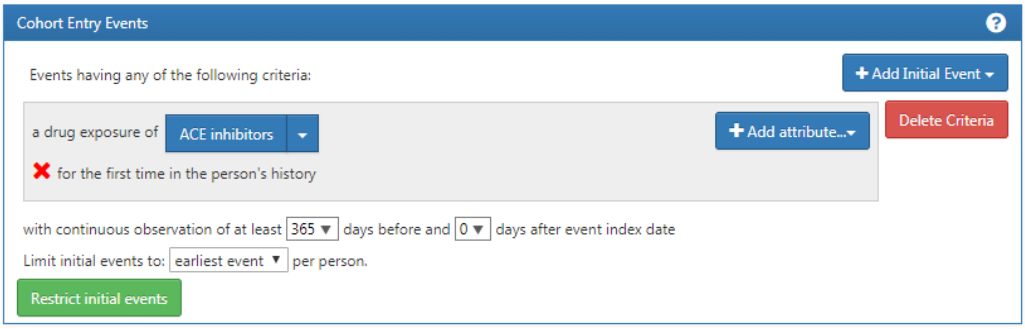

Cohort Entry Events
Events having any of the following criteria:
+ Add Initial Event
a drug exposure of
ACE inhibitors
+ Add attribute...
Delete Criteria
for the first time in the person's history
with continuous observation of at least
365
days before and
0
days after event index date
Limit initial events to:
earliest event
per person.
Restrict initial events

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

□□□□□□□□□□□□□□□□□□□

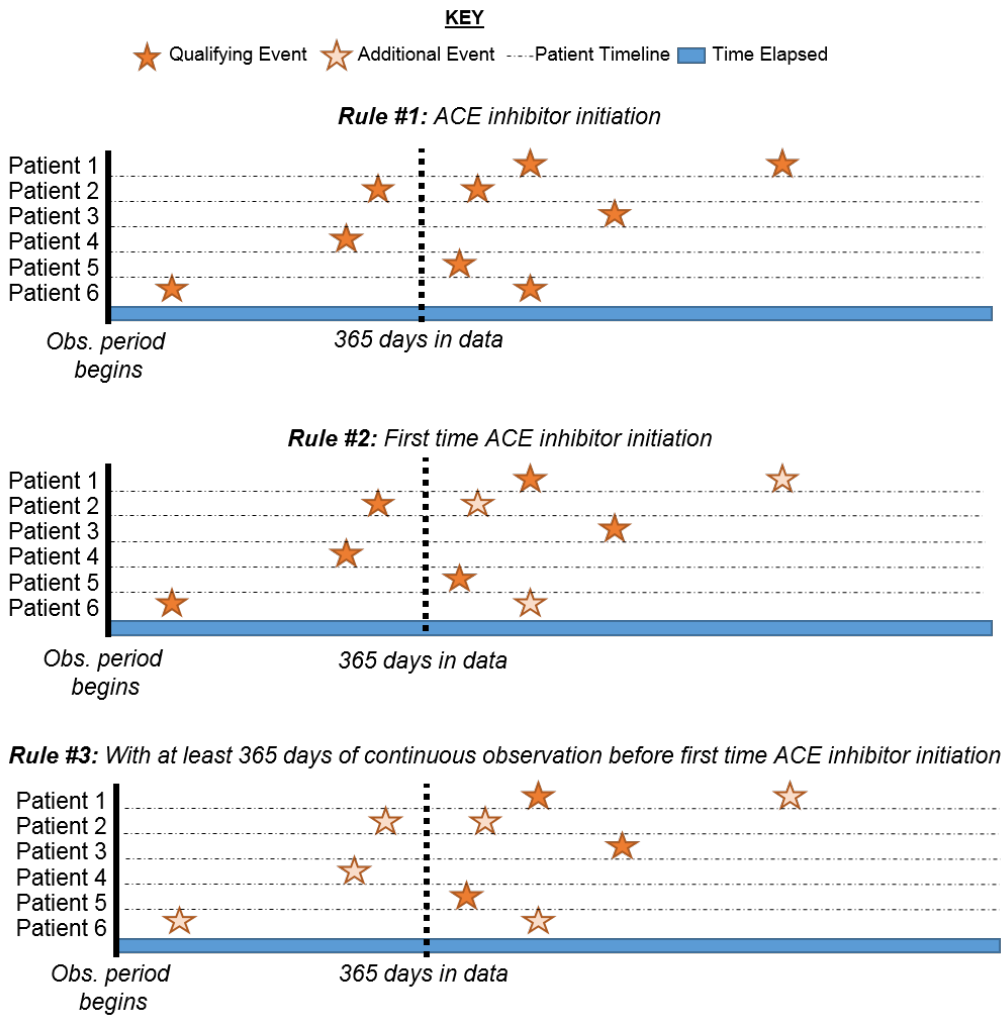

KEY
Qualifying Event
Additional Event
Patient Timeline
Time Elapsed
Rule #1: ACE inhibitor initiation
Patient 1
Patient 2
Patient 3
Patient 4
Patient 5
Patient 6
Obs. period begins
365 days in data
Rule #2: First time ACE inhibitor initiation
Patient 1
Patient 2
Patient 3
Patient 4
Patient 5
Patient 6
Obs. period begins
365 days in data
Rule #3: With at least 365 days of continuous observation before first time ACE inhibitor initiation
Patient 1
Patient 2
Patient 3
Patient 4
Patient 5
Patient 6
Obs. period begins
365 days in data

Figure 10.11: □□□□□□□□□□□□□□□□□□□□

```
                                        password = "supersecret")

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

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

```
                          sql,
                          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

□□□□□□□

```
          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□□□□□□□□□□□□□□□□□□□

□□□□□□□ACE□□□□□□□

10.9

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

- □□□□□ER□□□□□□□□□□□9201□9203□262□□□□□□□□□□□ □□□□□□□□□□□□

Chapter 11

11 – –

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

□□□□

Elm et al., 2008)□

11.3

□□□□□

$$= \frac{\#}{\quad}$$

□□□□□□□□□□□

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□□□□□□: □□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□

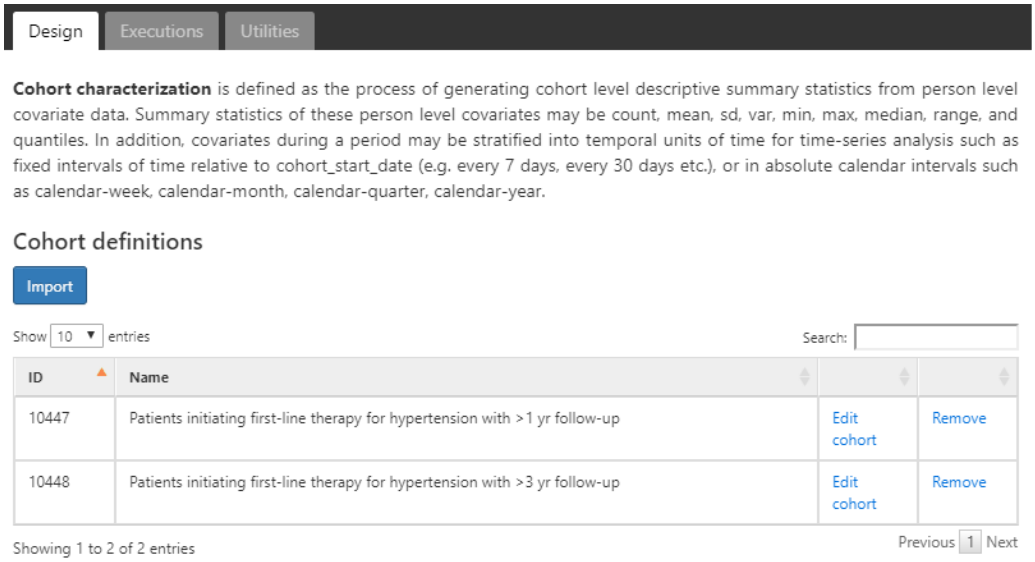


Figure 11.6: □□□□□□□ - □□□□□□□□□□□

□□□□□□

□□□□□□□□ATLAS□□□□□□□□□□□□□□□□□□□□□Chapter 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: □□□□□□□□ - □□□□□□□□□□□

ATLAS□□□□□

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: □□□□□□□□ - □□□□□□□□

□□□□explore□□□□□

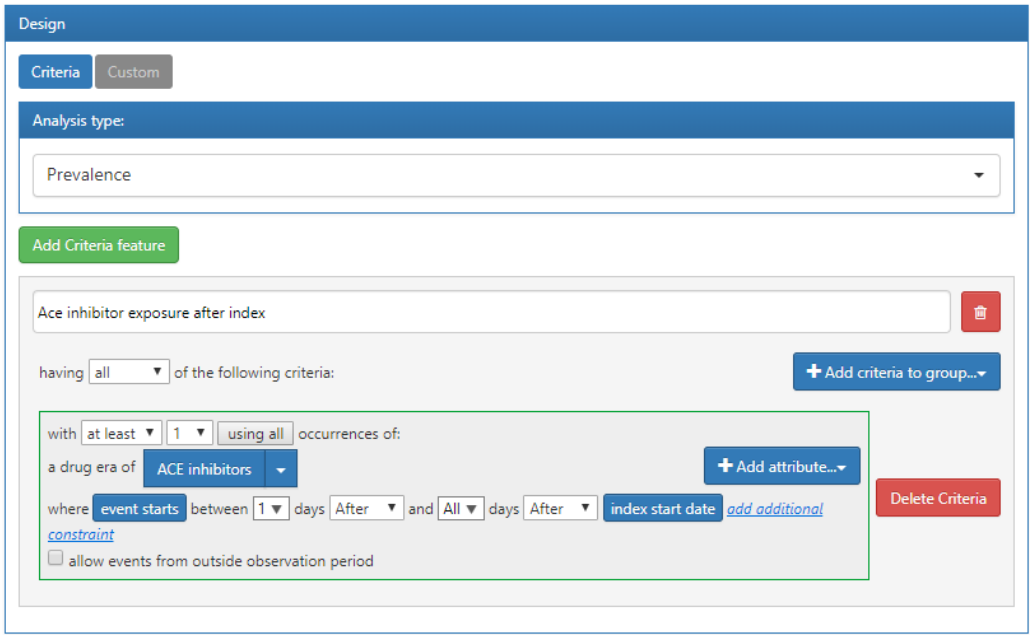
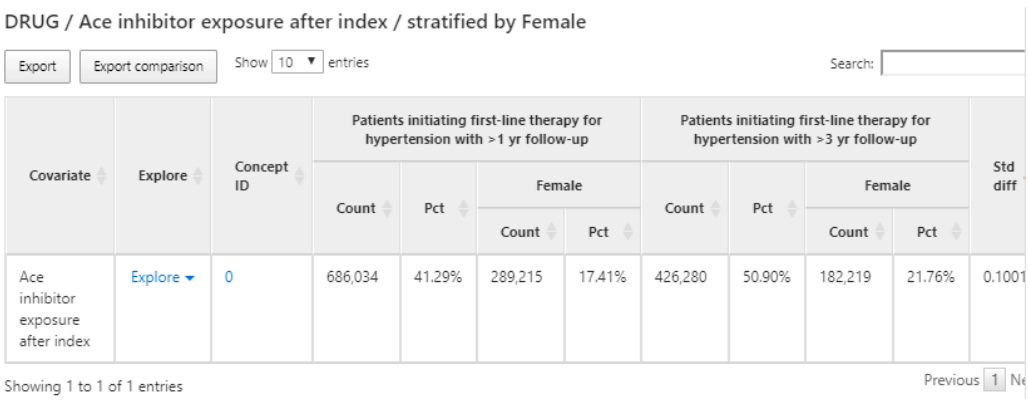


Figure 11.15: ATLAS□□□□□□□□□□□



DRUG / Ace inhibitor exposure after index / stratified by Female

Export | Export comparison | Show 10 entries | Search:

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	
Ace inhibitor exposure after index	Explore	0	686,034	41.29%	289,215	17.41%	426,280	50.90%	182,219	21.76%	0.1001

Showing 1 to 1 of 1 entries | Previous 1 Ne

Figure 11.16: □□□□□□□□□□□□□□□□□

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

FeatureExtraction□□□□□□□□□□□□□□□□□□□□□□□2□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

```
settings <- createCovariateSettings(useConditionEraLongTerm = TRUE,
                                    useConditionEraShortTerm = TRUE,
                                    useDrugEraLongTerm = TRUE,
                                    useDrugEraShortTerm = TRUE,
                                    longTermStartDays = -180,
                                    shortTermStartDays = -14,
                                    endDays = -1)
```

□□□□□□□□□□□□□□□□□□□□□□□□□□180□□□□□□□□□□□□□□□□□□□□□□□□□□□□14□□□

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

```
settings <- createCovariateSettings(useConditionEraLongTerm = TRUE,
                                    useConditionEraShortTerm = TRUE,
                                    useDrugEraLongTerm = TRUE,
                                    useDrugEraShortTerm = TRUE,
                                    longTermStartDays = -180,
                                    shortTermStartDays = -14,
                                    endDays = -1,
                                    excludedCovariateConceptIds = 1124300,
                                    addDescendantsToExclude = TRUE,
                                    aggregated = TRUE)
```

□□□□□□□□□□□□□□□□aggregated = TRUE□□□□□□□FeatureExtraction□□□□□□□□□□□

11.8.4

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

```
covariateSettings <- createDefaultCovariateSettings()

covariateData2 <- getDbCovariateData(
  connectionDetails = connectionDetails,
  cdmDatabaseSchema = cdmDatabaseSchema,
  cohortDatabaseSchema = resultsDatabaseSchema,
  cohortTable = "cohorts_of_interest",
  cohortId = 1,
  covariateSettings = covariateSettings,
  aggregated = TRUE)

summary(covariateData2)
```

□□□□□□□□□□□□□□□

```
## CovariateData Object Summary
##

## Number of Covariates: 41330

## Number of Non-Zero Covariate Values: 41330
```

11.8.5

□□□□□covariateData□□□□□

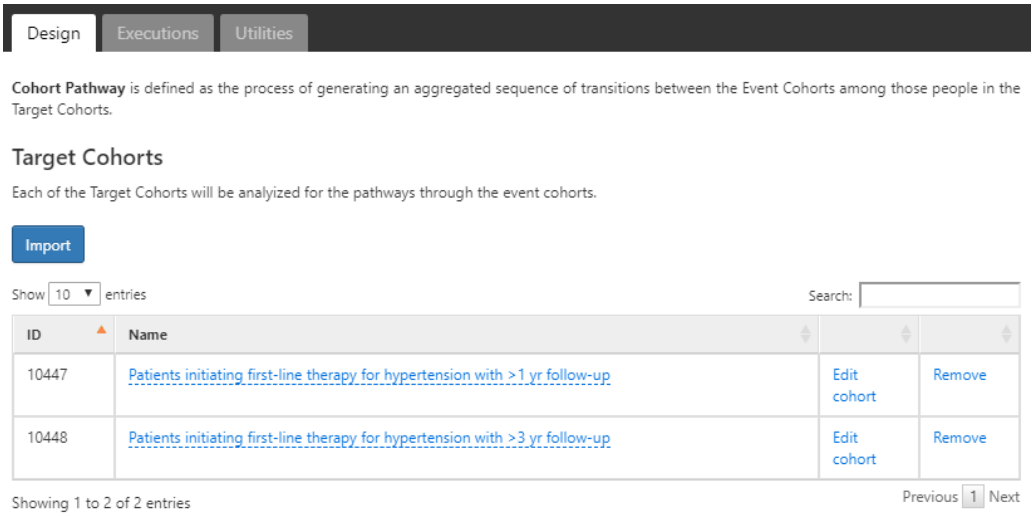

Design
Executions
Utilities
Cohort Pathway is defined as the process of generating an aggregated sequence of transitions between the Event Cohorts among those people in the Target Cohorts.
Target Cohorts
Each of the Target Cohorts will be analyized for the pathways through the event cohorts.
Import
Show 10 entries
Search:
ID
Name
10447
Patients initiating first-line therapy for hypertension with >1 yr follow-up
Edit cohort
Remove
10448
Patients initiating first-line therapy for hypertension with >3 yr follow-up
Edit cohort
Remove
Showing 1 to 2 of 2 entries
Previous 1 Next

Figure 11.18: □□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□

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: □□□□□□

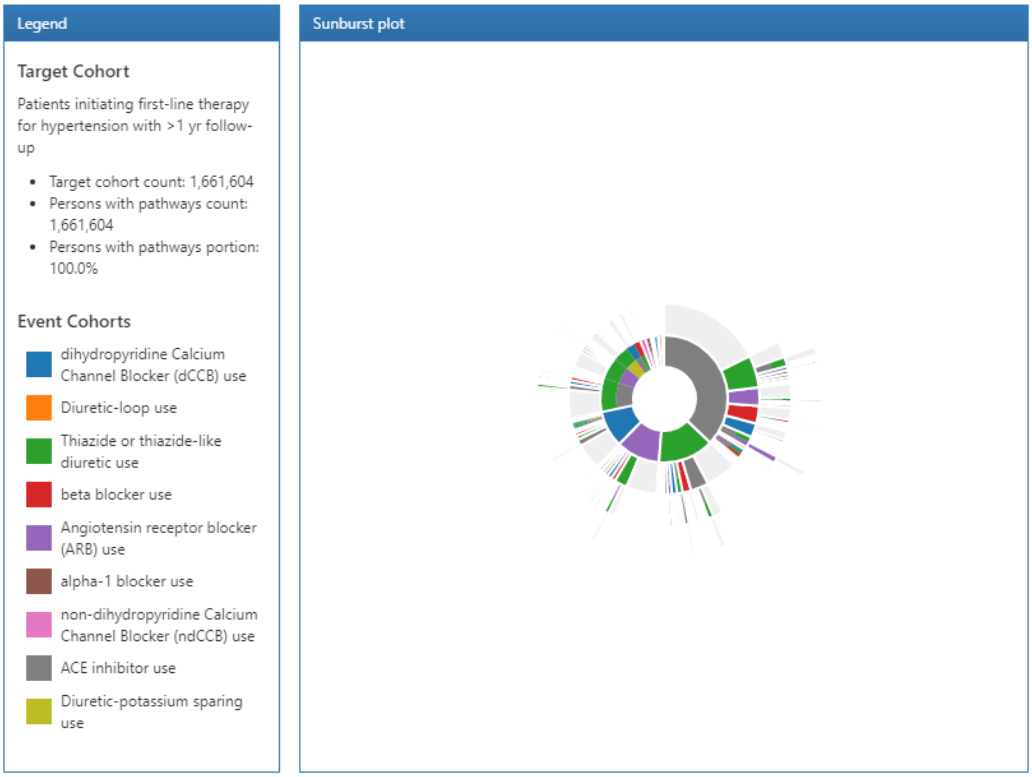

Legend
Target Cohort
Patients initiating first-line therapy for hypertension with >1 yr follow-up
Target cohort count: 1,661,604
Persons with pathways count: 1,661,604
Persons with pathways portion: 100.0%
Event Cohorts
dihydropyridine Calcium Channel Blocker (dCCB) use
Diuretic-loop use
Thiazide or 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: □□□□□□□□□□

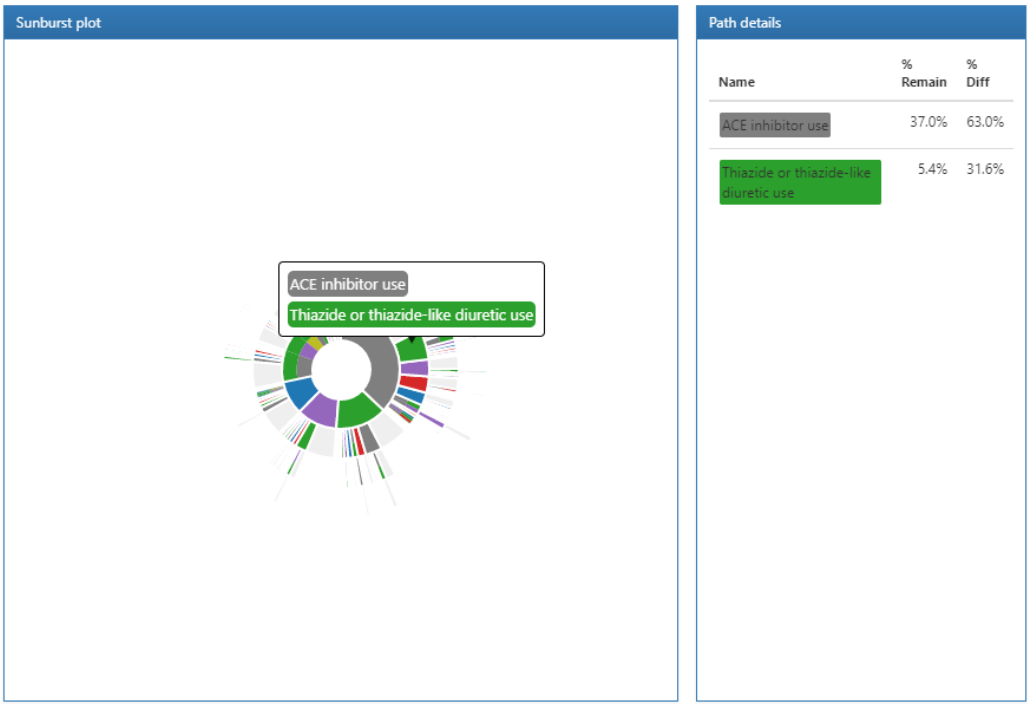


Figure 11.21: □□□□□□□□□□□□□□□□□□□□□□

(ARB) □□□□□□□□□□□□□□□□ARB □□□□□□□□□□□□□□□□□□□□

acute myocardial infarction events □ angioedema events □□□□ Angiotensin receptor blocker (ARB) use □□□□□□□□□□

Figure 11.25: □□□□□□□□

	Persons	Cases	Proportion [+|-] per 1k persons	Time At Risk (years)	Rate [+|-] per 1k years
Summary Statistics:	816,867	1,703	2.08	623,402	2.73

	Stratify Rule	N	Cases	Proportion [+|-] per 1k persons	Time At Risk (years)	Rate [+|-] per 1k years
1.	Gender = Female	313,518	416	1.33	230,157	1.81

Figure 11.26: □□□□□□□□ - AMI □□□□□□□□□□□□ ACEi □□□□

□□□

□□ **11.1.** □□□□□□

Chapter 12

12 – –

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

□□□□□□□□

□: □□□□□□□□□□□□□□□□□□□□

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

□□□□□□□□□□□□□

Table 12.2: □□□□□□□□□□□□□□□□□□□□□□□□

Choice	Description
Target cohort	□□□□□□□□□□
Outcome cohort	□□□□□□□□□□□□□□□□□□
Time-at-risk	□□□□□□□□□□□□□□□□□□□□

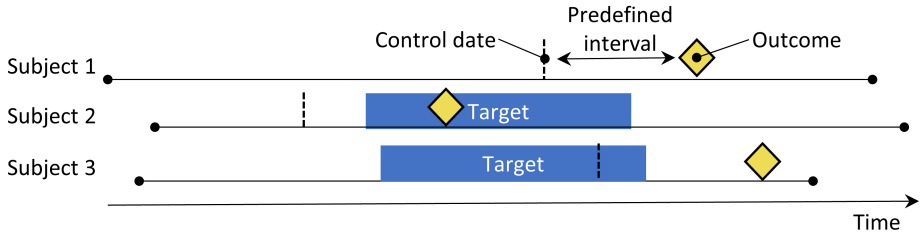

Predefined
interval
Control date
Outcome
Subject 1
Subject 2
Target
Subject 3
Target
Time

Figure 12.4: □□□□□□□□□□□□

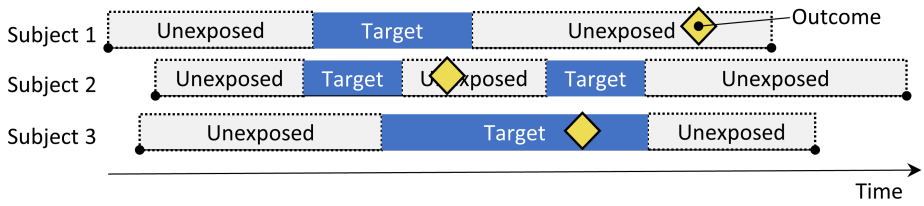

Subject 1
Unexposed
Target
Unexposed
Outcome
Subject 2
Unexposed
Target
Target
Unexposed
Subject 3
Unexposed
Target
Unexposed
Time

Figure 12.5: □□□□□□□□□

12.6.7

□□□□□□□□□□□

12.7□□□□□□□□

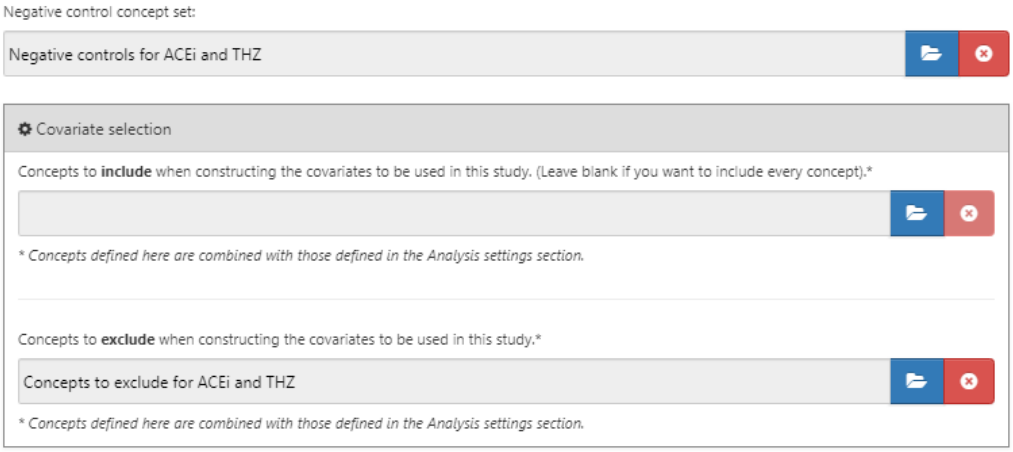

Negative control concept set:
Negative controls for ACEi and THZ
Covariate selection
Concepts to **include** when constructing the covariates to be used in this study. (Leave blank if you want to include every concept).*
* Concepts defined here are combined with those defined in the Analysis settings 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: □□□□□□□□□□□□

```
library(CohortMethod)
connDetails <- createConnectionDetails(dbms = "postgresql",
                                       server = "localhost/ohdsi",
                                       user = "joe",
                                       password = "supersecret")

cdmDbSchema <- "my_cdm_data"
cohortDbSchema <- "scratch"
cohortTable <- "my_cohorts"
cdmVersion <- "5"
```

□□□4□□cdmDbSchema□cohortDbSchema□□□□cohortTable□□□□CDM□□□□□□□□□□□□□
SQL Server□□□□□□□□□□□□□

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

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

□□□□□1□□□□□□□□□□□□□□□□CohortMethod□□□□□□□□□□□□□□□□

12.8.5

□□□□□□□□□□□

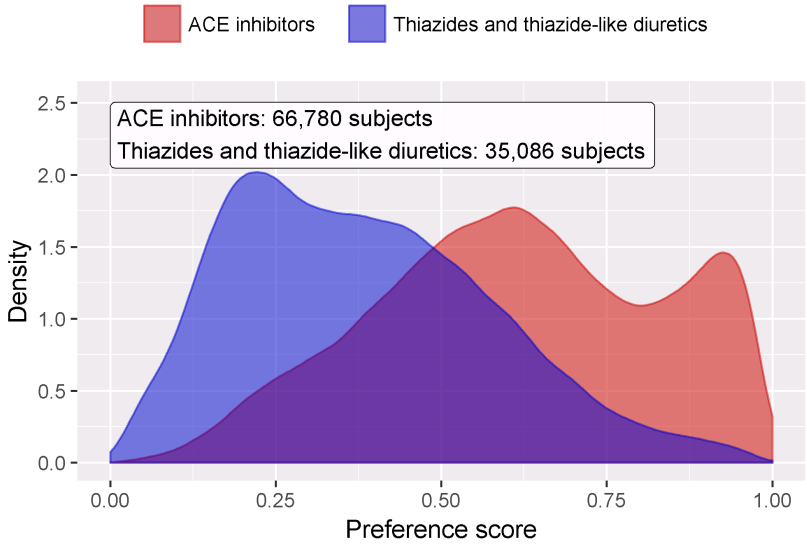


Figure 12.10: □□□□□□□□

□□□	□□□
-1.42	□□□□□-30□□□0□□□□□□□□□□□□: □□
-1.11	□□□□□0□□□0□□□□□□□□□□□□: □□□□□□
0.68	□□□□□□: 05-09
0.64	□□□□□-365□□□0□□□□□□□□□□: □□□
0.63	□□□□□-30□□□0□□□□□□□□□□□□: □□□
0.57	□□□□□-30□□□0□□□□□□□□□□□□: □□□□□
0.55	□□□□□-365□□□0□□□□□□□□□□□□: □□□□□□□□□□□
-0.54	□□: □□□□□□□□□□□□□□□□
0.52	(□□)
0.50	□□: □□

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

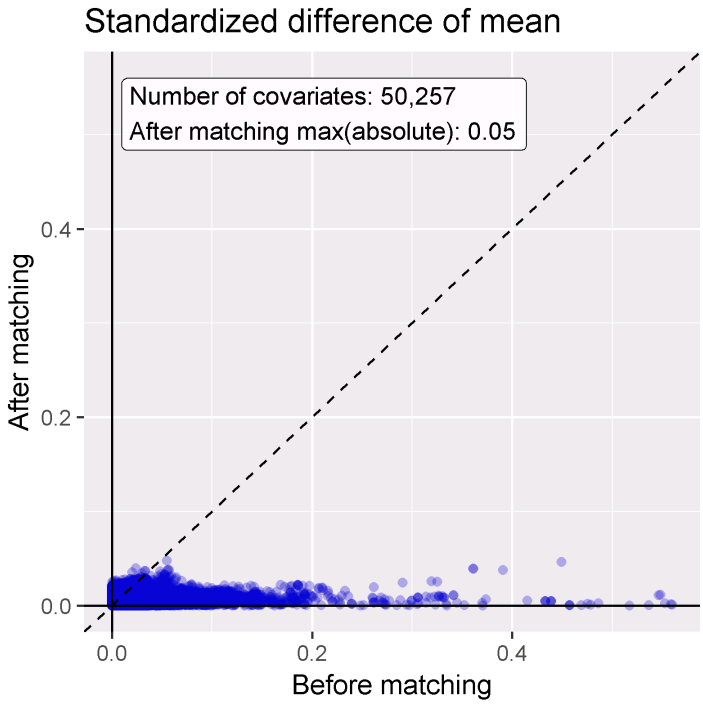

Standardized difference of mean
Number of covariates: 50,257
After matching max(absolute): 0.05
After matching
0.0
0.2
0.4
Before matching
0.0
0.2
0.4

Figure 12.11: □□□□□□□□□□□□□□□□□ □□□□□□□□□□□□□□□□□

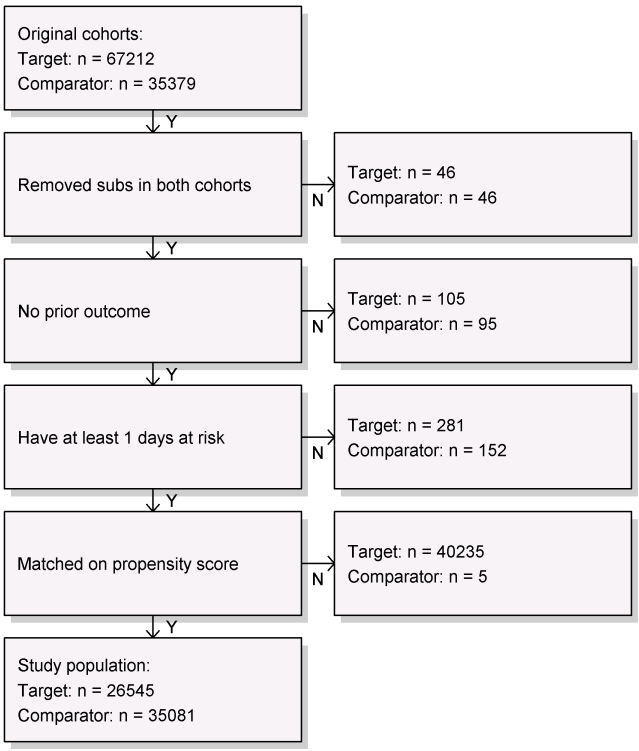

Original cohorts:
Target: n = 67212
Comparator: n = 35379
Y
Removed subs in both cohorts
N
Target: n = 46
Comparator: n = 46
Y
No prior outcome
N
Target: n = 105
Comparator: n = 95
Y
Have at least 1 days at risk
N
Target: n = 281
Comparator: n = 152
Y
Matched on propensity score
N
Target: n = 40235
Comparator: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

Figure 12.12: □□□□□□□□□

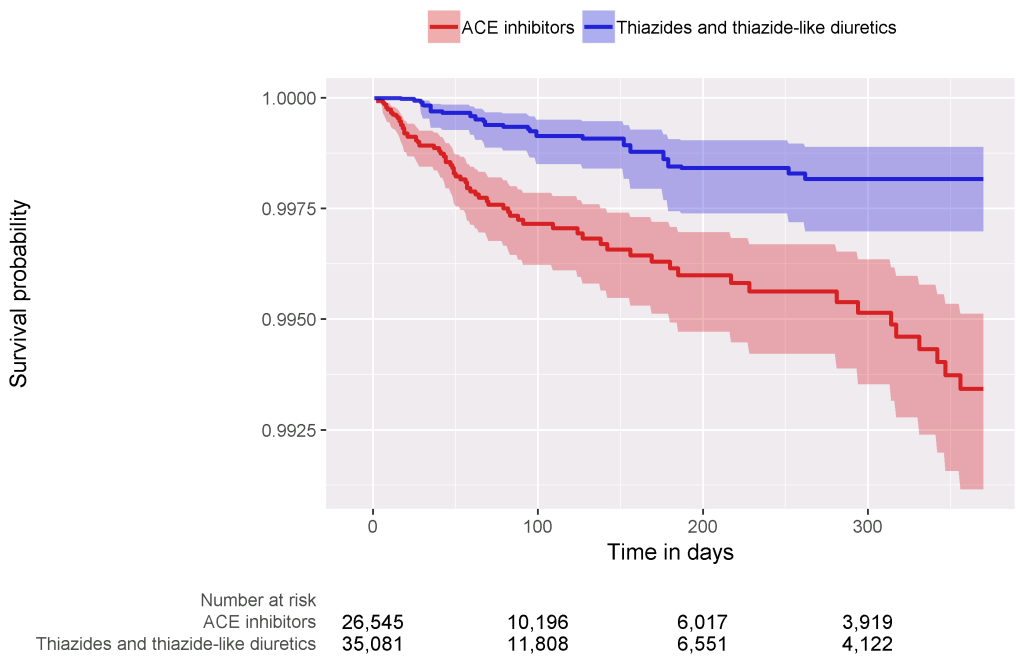


Figure 12.14: □□□□−□□□□□□□□□□

```
remotes::install_github("ohdsi/CohortMethod")
```

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

□□ **12.3.** □□□□□□□□□

Chapter 13

13 – –

□*Lead: Peter Rijnbeek & Jenna Reps*

□□□

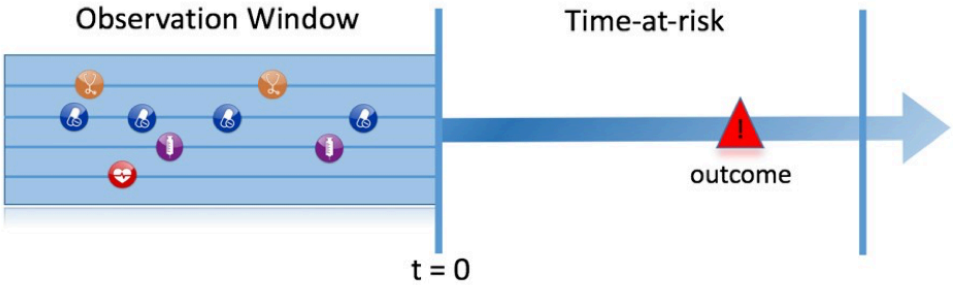


Figure 13.1: □□□□□

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

13.2.1

□ 13.2.1 □□2□□□□□□□□□□□□COHORT□□□□□□□□□□□□□□□□□□□□□□□□ID□1□□□□□

□□ □□□□COHORT□□□□□□□□□□□□COHORT_END_DATE□□□□□□□□□□

COHORT_DEFINITION_ID	SUBJECT_ID	COHORT_START_DATE
1	1	2000-06-01
1	2	2001-06-01
2	2	2001-07-01

□ 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

□□□□□□□□□□□□□□□□□□□

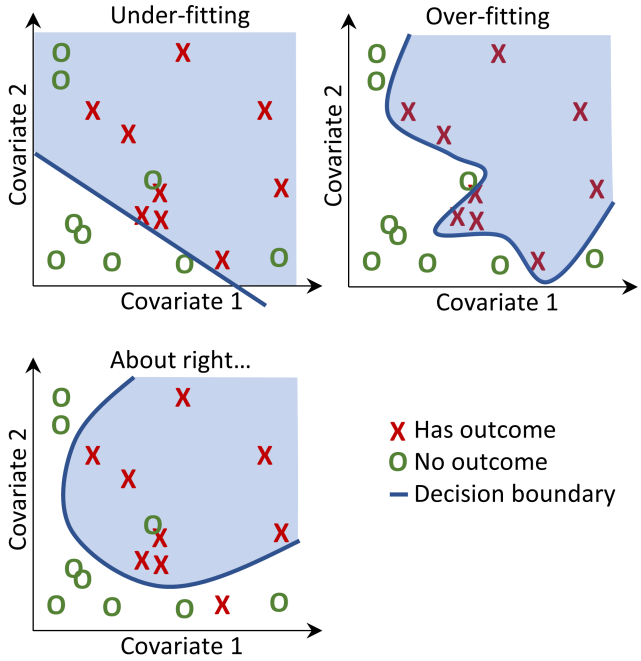


Figure 13.2: Decision boundary.

13.3.1

LASSO□□□□□□□

□□□□□	□□	□□□□□
learningRate	□□□□□□□□□□□□	0.005,0.01,0.1
maxDepth	□□□□□□	4,6,17
minRows	□□□□□□□□□□□□□□□□	2
ntrees	□□□	100,1000

13.3.3

□□□□□□□□□□□□□□□

□□□□□	□□	□□□□□

13.3.7

□□□□□□□

□□□□□□□□

- □□: $TP/(TP+FN)$
- □□□: $TN/(TN+FP)$
- □□□□□: $TP/(TP+FP)$

□□□□□□□

13.5

□□□□□□□□□□□□□

13.5.5

□□□□□ 13.5.5□□□□□□□□□□□□□□□□□□□□□□□□□□

□: □□□□□□□□□□□□□□□

□□	□
□□□□□□□□□□ □□□□□□□□□□ □□□□□ □□□	□□□ACE□□□□□□□□□□□□□□□□□□□□□□365□□□□□□□□□□□□□□□□□□□

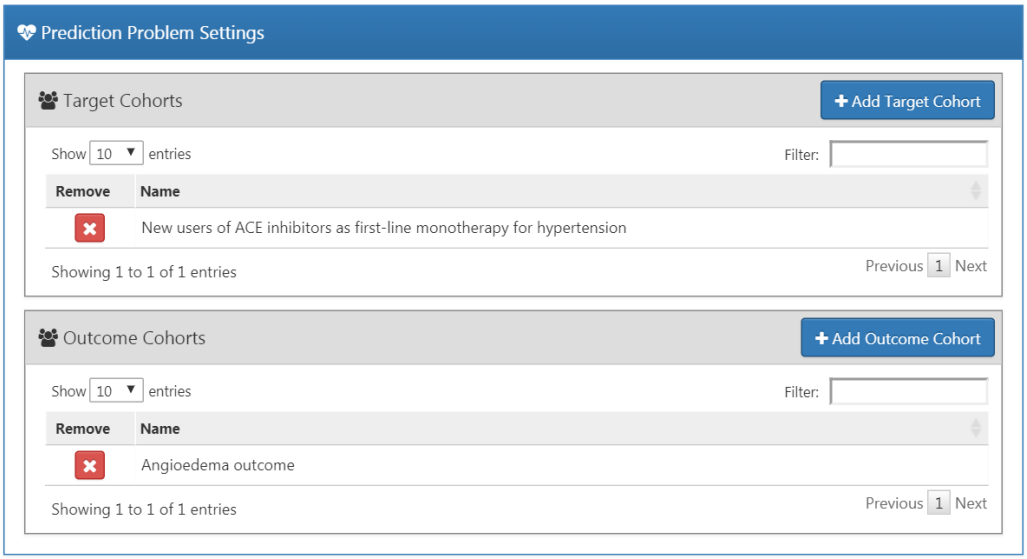


Figure 13.4: □□□□□□.

□□□□□

CDM□□□□□□□□

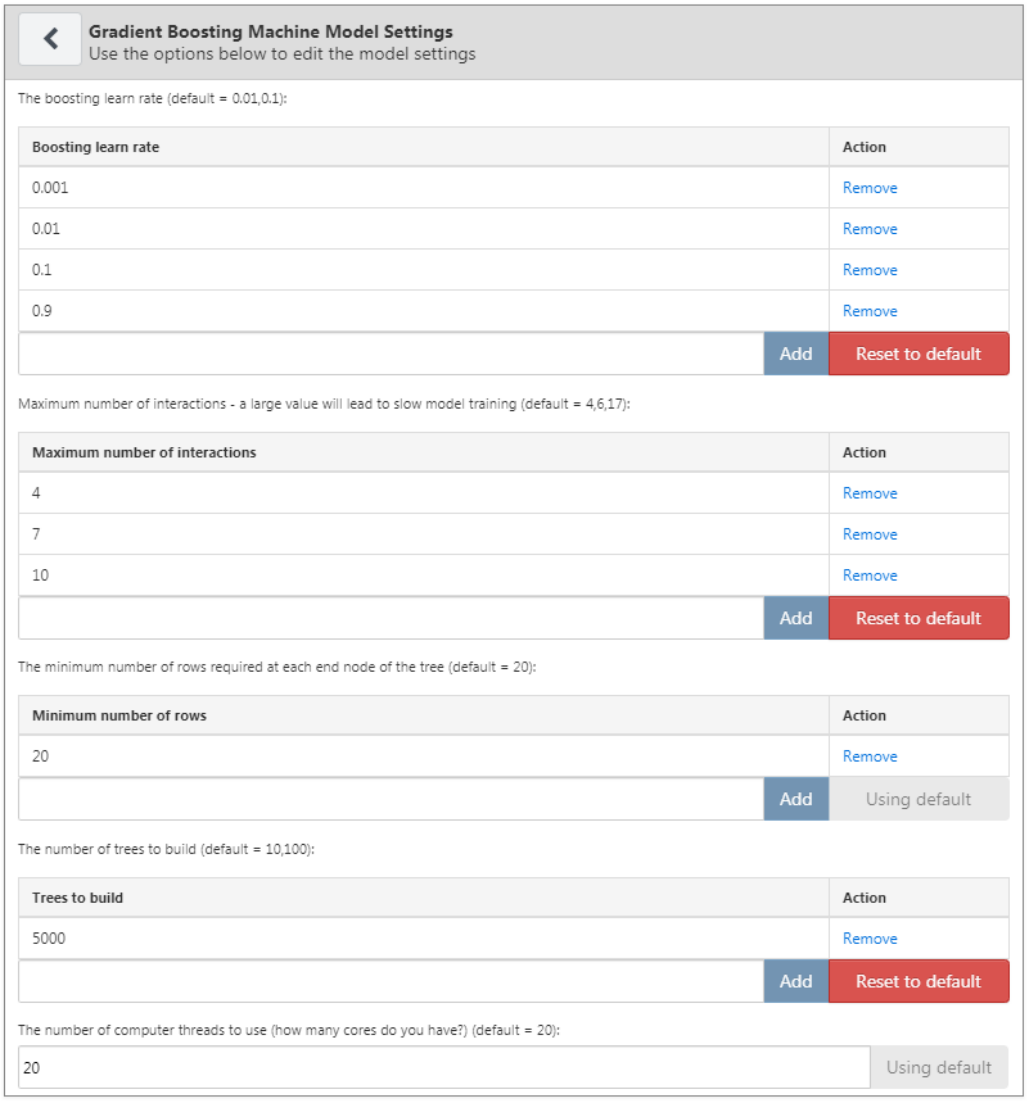

Gradient Boosting Machine Model Settings
Use the options below to edit the model settings
The boosting learn rate (default = 0.01,0.1):
Boosting learn rate
Action
0.001
Remove
0.01
Remove
0.1
Remove
0.9
Remove
Add
Reset to default
Maximum number of interactions - a large value will lead to slow model 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: □□□□□□□□□□□□.

□□□□□□□□□□□□□□□□□□□□ 13.7□□□□□□□□□□□□

Select Covariates

	Gender	Age	Age Groups	Race	Ethnicity	Index Year	Index Month	Prior Observation Time	Post Observation Time	Time In Cohort	Index Year & Month
Demographics	☑	☐	☑	☑	☑	☐	☑	☐	☐	☐	☐

Figure 13.7: □□□□□□.

□□□□□□□□□□□□□□3□□□□□□□□□□□□□□□

- □□□: □□□□□□□□□□□□□□[□□□□□□□0□]
- □□ [□□□□□□-365□□□□□□□□□□□□□]
- □□ [□□□□□□-180□□□□□□□□□□□□□]
- □□ [□□□□□□-30□□□□□□□□□□□□□]

□□□□□□□□□□□□□□□□□□□□ 13.8□□□□□□□□□□□□

Time bound covariates

Set the time windows for the time bound covariates in days relative to the cohort index

	Any Time Prior	Long Term	Medium Term	Short Term	End Days
Time Windows	All Time	-365	-180	-30	0

Figure 13.8: □□□□□□□□□□□.

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

- Condition: □□□□□□□□□□□□□□□□□□□□□□□□□□ID□□□□□□□□□□□□□□□□□□□

□□□□□□□

```
cohortsDbTable <- "my_cohorts"
cdmVersion <- "5"
```

□□□4□□cdmDbSchema□cohortsDbSchema□cohortsDbTable□□□□□□□CDM□□□□□□□
SQL Server□□□□□□□

```
                    sampleSize = 10000
)
```

getPlpData□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□PatientLevelPrediction□□□□□□□□□□

plpData□□□□□□□□□□□□□□□□□□□

runPlP□□□□□□□plpData□□□75
25%□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□testSplit□□/□□□□□□□testFractio

```
gbmResults <- runPlp(population = population,
                     plpData = plpData,
                     modelSettings = gbmModel,
                     testSplit = 'person',
                     testFraction = 0.25,
                     nfold = 2,
                     splitSeed = 1234)
```

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

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

□□□□□□□□□□□□□

```
savePlpModel(gbmResults$model, dirPath = "model")
```

□□□□□□□□□□□□□□

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

□□□□□□□□□□□□□□□□□□□□□□□□□□□□

```
savePlpResult(gbmResults, location = "gbmResults")
```

□□□□□□□□□□□□□□□□□□□□□□□

```
gbmResults <- loadPlpResult("gbmResults")
```

13.7.5

□□□□□□□□□□□runPlp□□□□□□□□□□□□□□□□□□□□□□/□□□□□□□□□□□□□□□□□□□

13.7.6

□□□□□□□

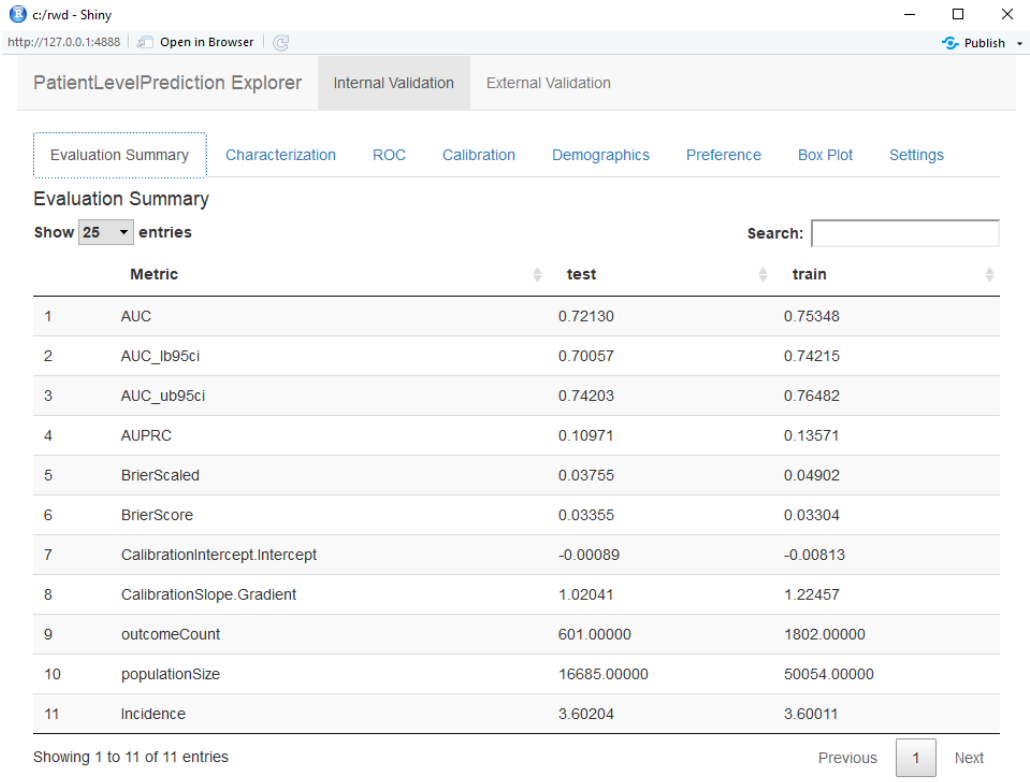


Figure 13.10: Shiny□□□□□□

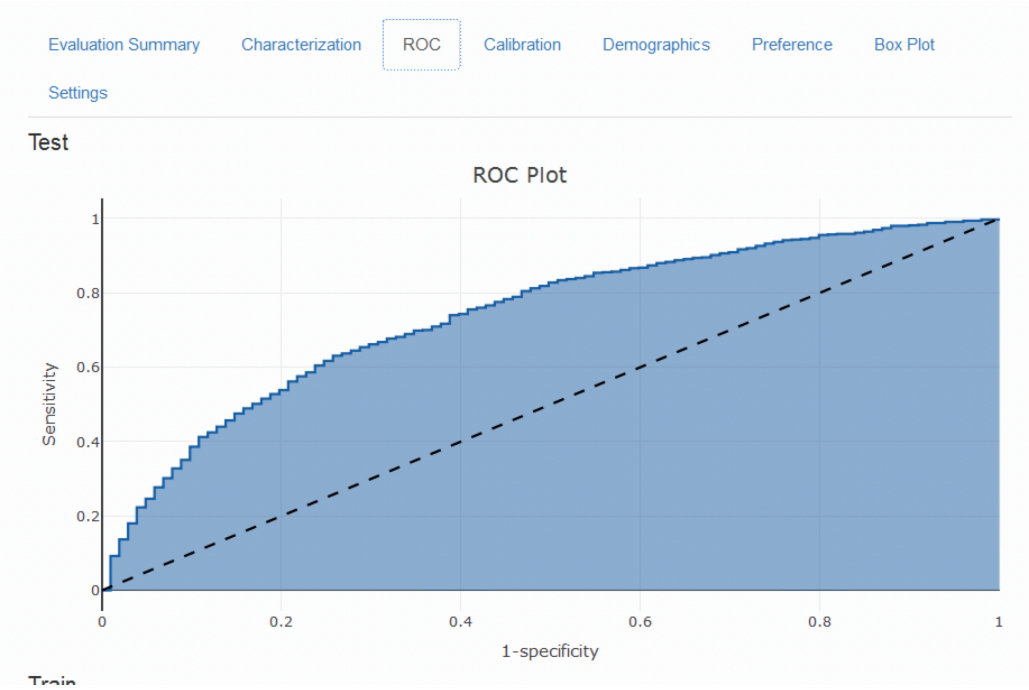


Figure 13.11: ROC□□□□□

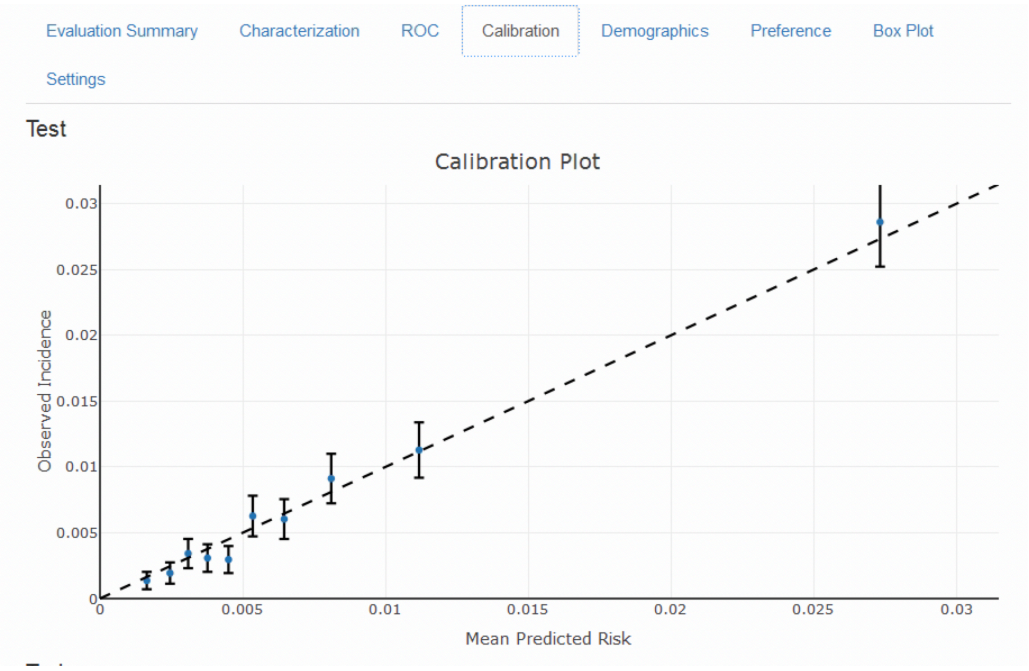


Figure 13.12: □□□□□□□□□□□□□□

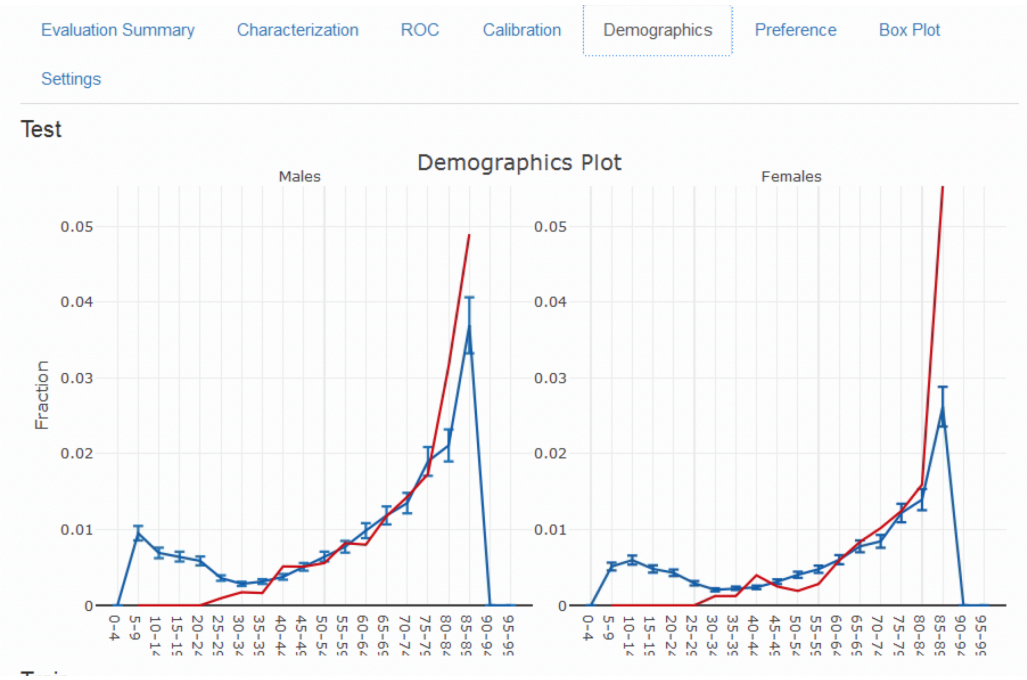


Figure 13.13: □□□□□□□□□□□□□□□□□□□□□□

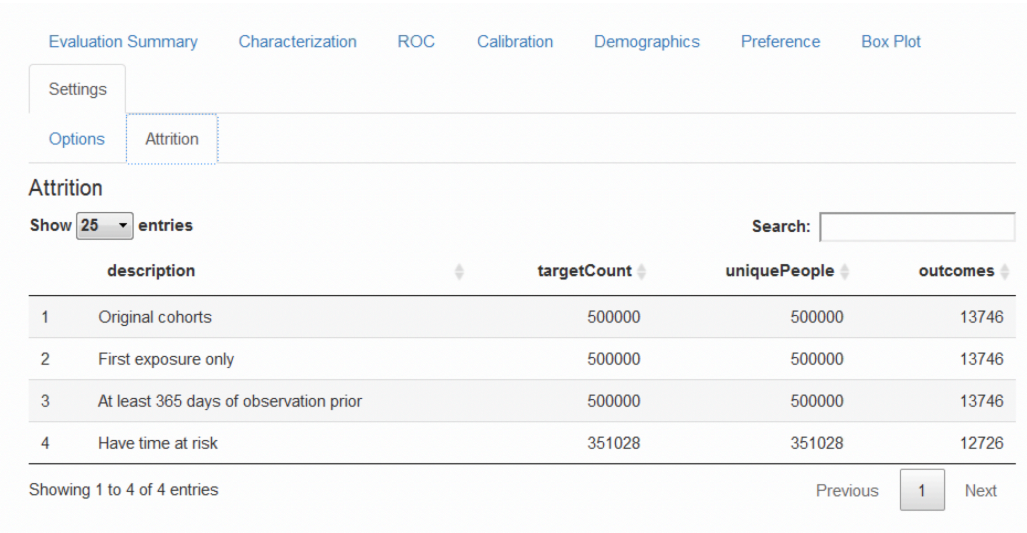


Figure 13.14: □□□□□□□□attrition□□□□

□□□□□□□□□□□□

□□□□□□□□□□Shiny□□□□□□ 13.15□□□□□□□□□□□□□□□□□□

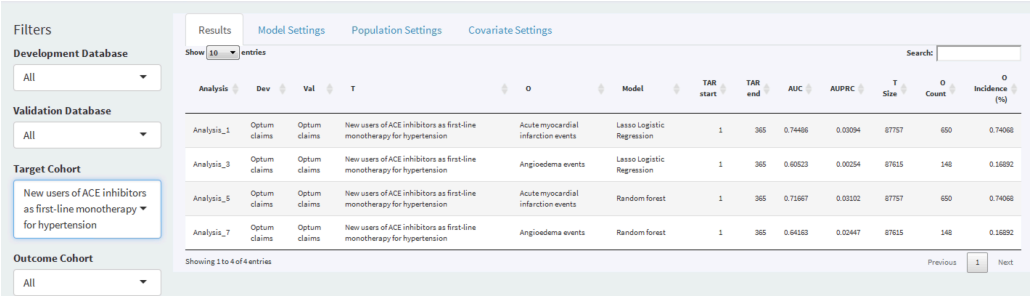


Figure 13.15: □□□Shiny□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□

- □□□□□□□□□□□□□□□□□□□□

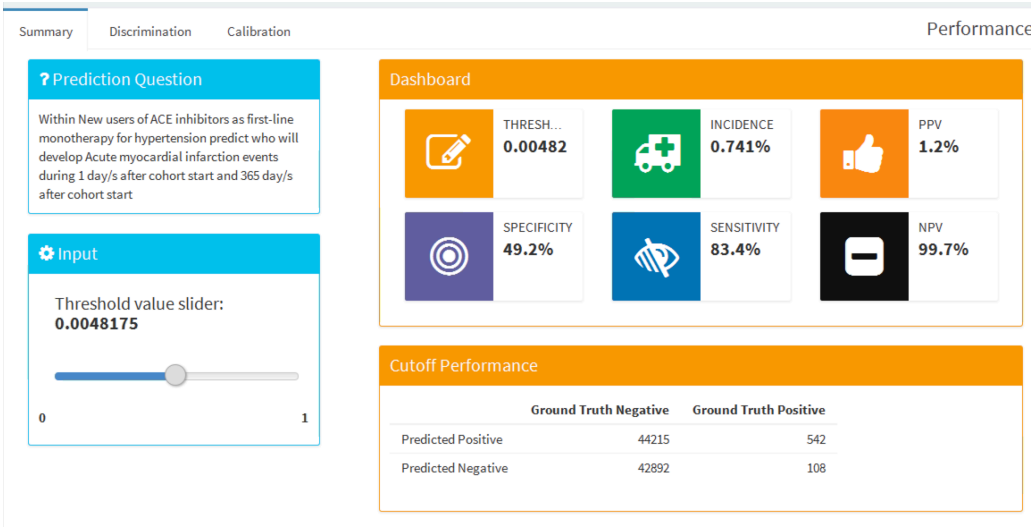

Summary
Discrimination
Calibration
Performance
Prediction Question
Within New users of ACE inhibitors as first-line monotherapy for hypertension predict who will develop Acute myocardial infarction events during 1 day/s after cohort start and 365 day/s after cohort start
Input
Threshold value slider:
0.0048175
0
1
Dashboard
THRESH...
0.00482
INCIDENCE
0.741%
PPV
1.2%
SPECIFICITY
49.2%
SENSITIVITY
83.4%
NPV
99.7%
Cutoff Performance
Ground Truth Negative
Ground Truth Positive
Predicted Positive 44215 542
Predicted Negative 42892 108

Figure 13.17: □□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□

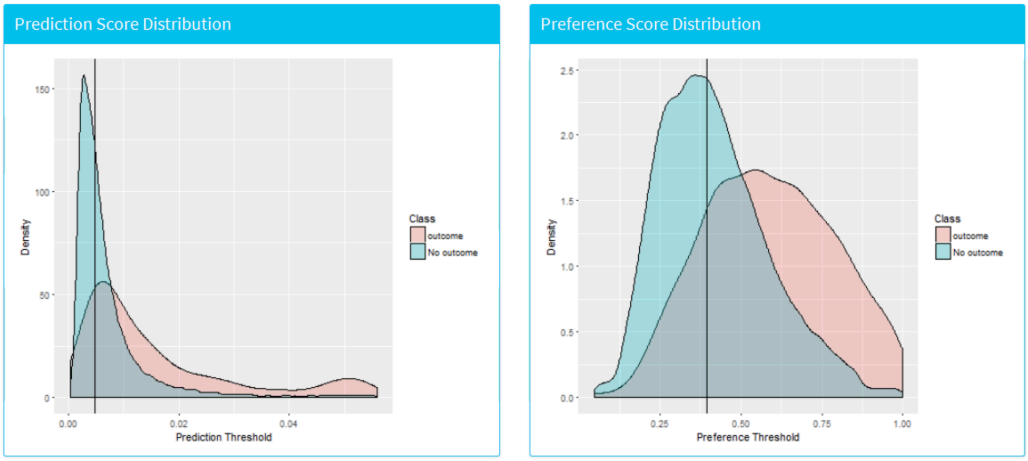


Figure 13.19: □□□□□□□□□□□outing□□□□□□□□□□

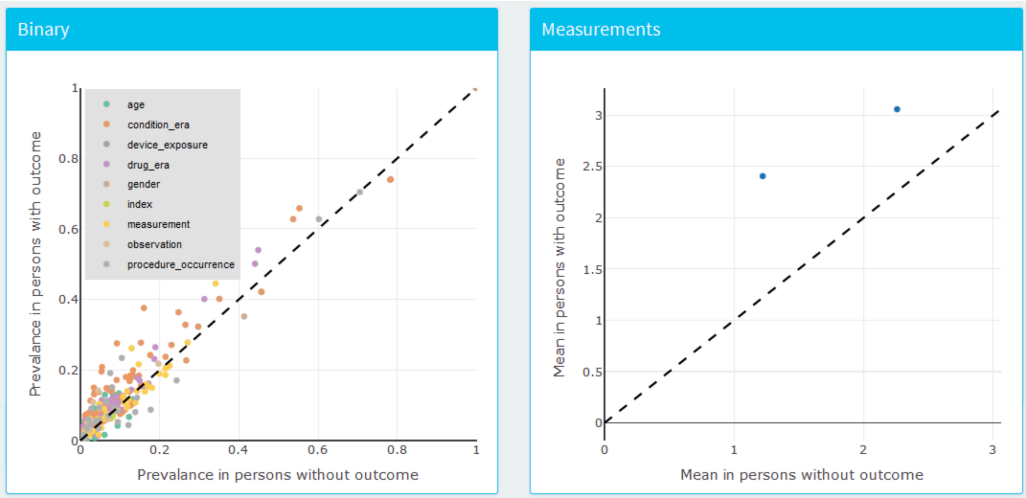

Binary
age
condition_era
device_exposure
drug_era
gender
index
measurement
observation
procedure_occurrence
Prevalance in persons with outcome
Prevalance in persons without outcome
Measurements
Mean in persons with outcome
Mean in persons without outcome

Figure 13.21: □□□□□□□

□□□□□□□

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

Eunomia□□□□□□□□CDM□□□□□□□□

Part II

Chapter 14

*: Patrick Ryan & Jon Duke ======= *: Patrick Ryan & Jon Duke
b6a32c0f87976a25fc93e3647d145223df6fe11b

14.1

Desired attribute	Question	Researcher	Data	Analysis		Result
Repeatable	Identical	Identical	Identical	Identical	=	Identical
Reproducible	Identical	Different	Identical	Identical	=	Identical
Replicable	Identical	Same or different	Similar	Identical	=	Similar
Generalizable	Identical	Same or different	Different	Identical	=	Similar
Robust	Identical	Same or different	Same or different	Different	=	Similar
Calibrated	Similar (controls)	Identical	Identical	Identical	=	Statistically consistent

Figure 14.1:

< HEAD
Hill 1965
=======
Austin Bradford Hill
1965)
et al. (2013b)
b6a32c0f87976a25fc93e3647d145223df6fe11b

□□□□□□□

14.3

«««<HEAD □□□□□□□□□□□□

Chapter 15

— —

□□□□: *Martijn Schuemie, Vojtech Huser & Clair Blacketer*

□□□□

□□□□3□□□□□□□□□□□□□□□□□□□□□□□□OHDSI□□□□□□□□□□□□□□□□□□□□□□□□DQ□□□□
16□Chapter 17□Chapter 18 □□□□□□□□□□□□□ ## □□□□□□□□□□

□□□Kahn et al.
(2016) □□□□□□□□□□□□□□□□□□□□DQ□□□□3□□□□□□□□□□□□□□□□□□□□□□

1. □□□: □□□□□□□□□□□□□□□□□□□□□□□□□□3□□□□□□□□□□□□□□□□
   - □: □□□□□□□□□□□□□□□□□□□□

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

□□□□□□□□□□□□□□□□□□□

ETL □□□□□□

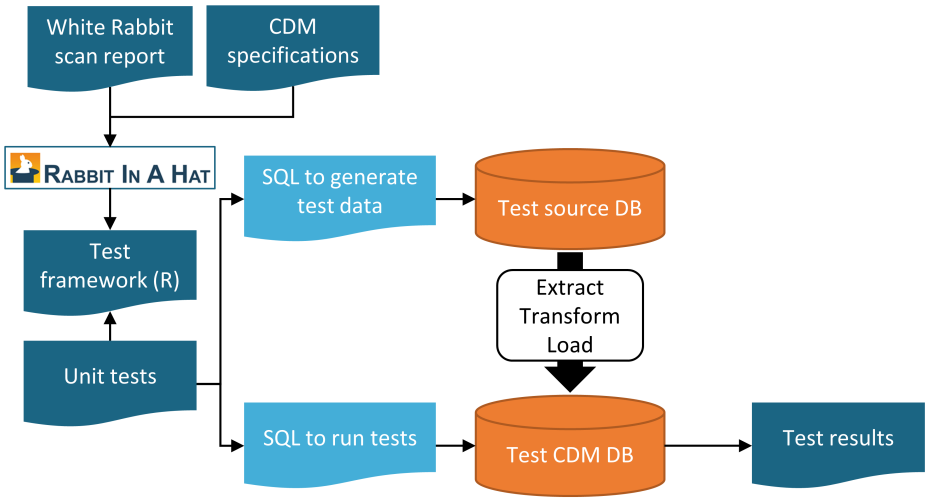


Figure 15.1: Rabbit-in-a-Hat □□□□□□□□□□□□□□□□□ ETL (Extract-Transform-Load) □□□□□□□□□□□

□□□□□□□□□□□ACHILLES□□□□□□□□□□□□□□□□

‰ per month	Max monthly ‰	Person count	Description
	26.81	92,019,885	**Depressive Disorder**
	6.64	15,969,198	**Depressive disorder** 440383
	6.64	15,686,275	311 (ICD9CM) Depressive disorder, not elsewhere classified
	0.46	188,230	F328 (ICD10CM) Other depressive episodes
	0.38	94,693	F3289 (ICD10CM) Other specified depressive episodes
	3.10	12,010,783	**Adjustment disorder with mixed emotional features** 433454
	3.07	9,839,712	30928 (ICD9CM) Adjustment disorder with mixed anxiety and depressed mood
	3.03	2,049,618	F4323 (ICD10CM) Adjustment disorder with mixed anxiety and depressed mood
	0.04	121,453	3091 (ICD9CM) Prolonged depressive reaction
	3.17	9,237,192	**Dysthymia** 433440

Figure 15.2: checkCohortSourceCodes□□□□□□□□□□

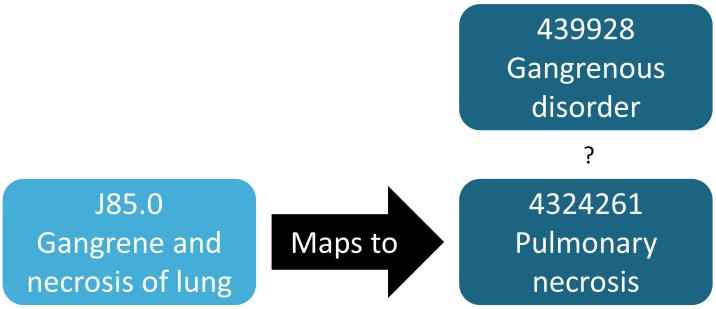


Figure 15.3: □□□□□□□□□□□□□□□□□

15.3 ACHILLES

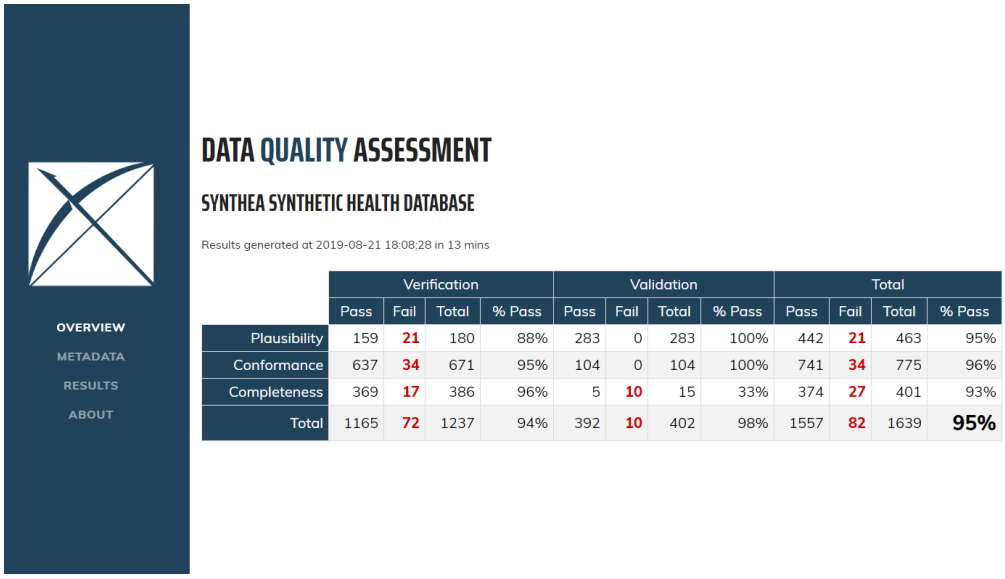
□□□□□CDM□□□□□□□□□□□□□□□ACHILLES□□□□□□□□□□□□□□□

□□□R□□□□□□□□□□□□□□□□□□□□□□□□□□□ACHILLES□ DatabaseConnector □□□□□□□□□□□□□□□□□□□□□ `createConnectionDetails` □□□□□□□□□□□□□□□□□□□□□

```
library(Achilles)
connDetails <- createConnectionDetails(dbms = "postgresql",
                                       server = "localhost/ohdsi",
                                       user = "joe",
                                       password = "supersecret")

cdmDbSchema <- "my_cdm_data"
cdmVersion <- "5.3.0"
```

□□□2□□ `cdmDbSchema` □□□CDM□□□□□□□□□□□□□□□□□□□□□□□CDM□□□□□□□□□□ SQL Server□□□□□□□□□□□□□□□□□□□



	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	159	21	180	88%	283	0	283	100%	442	21	463	95%
Conformance	637	34	671	95%	104	0	104	100%	741	34	775	96%
Completeness	369	17	386	96%	5	10	15	33%	374	27	401	93%
Total	1165	72	1237	94%	392	10	402	98%	1557	82	1639	95%

Figure 15.4: Data Quality Dashboard□□□□□□□□□□□□□□□□□□

□□□□□□□□□*Results*□□15.5□□□□
□□□□□□□□□□□□□□□□□□□

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 15.5: Data Quality Dashboard□□□□□□□□□□□□□□□□□□□

‰ 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 15.6: □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

```
                                                    "Giant hives",
                                                    "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

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□”Angioneurotic edema, sequela”□□□□□□□□□□□□□□□□□□□

```
remotes::install_github("ohdsi/DataQualityDashboard")
remotes::install_github("ohdsi/Eunomia", ref = "v1.0.0")
```

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

Chapter 16

— —

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

> □□□□□

16.2.1

□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□PPV□□□□1□□□□□□□□□□□□□□□□□□□

16.4.1 PheValuator

□□□□□□

□□□□ 2: xSens □□□□□□□

□□□□□□□□□□□□□□□□□xSens□□□□□□□□□□□□□□□□□

```
  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" □□□□□□□

```
  startDate = "20100101",
  endDate = "20171231")
```

□□□□□□□□□□□□□□□□□□□□c:/temp/lr_results_5XMI_train_myCDM_ePPV0.75_20181206V1

□□□□ **5:** □□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□

Chapter 17

— —

□□□□□□□□: *Martijn Schuemie*

□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□

- □□□□□: □□□□□□□□□□□

17.2.2

□□□□□□□□□□□□□□□□□□□□□□□□R□□□□□□□□□□□□□□□□□□□

17.4

- □□□□□□□CDM□□□□□□□□□□

Chapter 18

— —

□□□□□: *Martijn Schuemie*

□□□□□

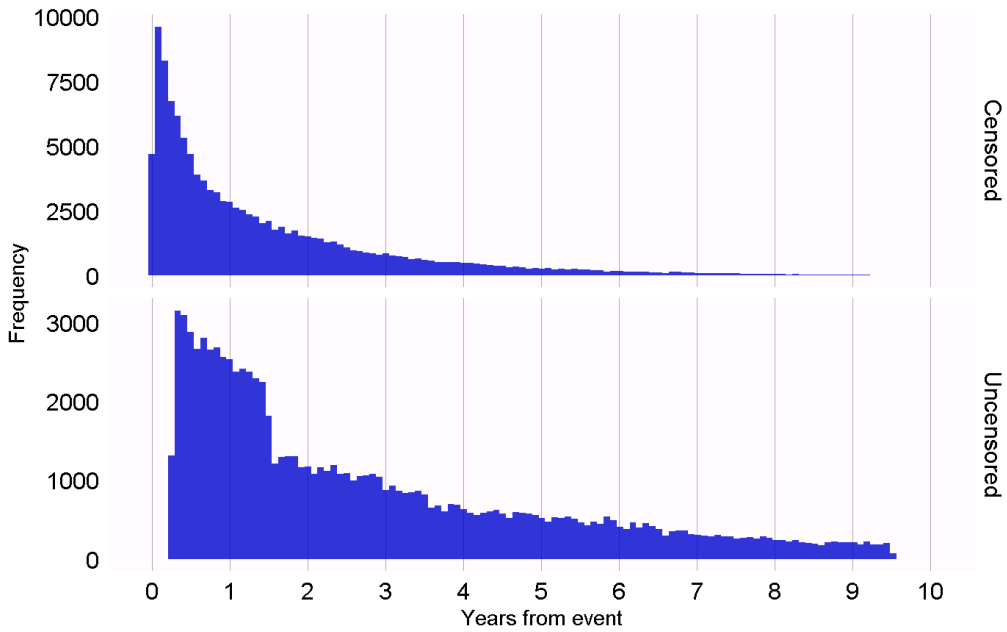


Figure 18.1: □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

18.2.1

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□-□□□□□□□□□□□□□□□□□(Lipsitch et al., 2010) □□□□□□□□□□□□□□□□□□□□

□□□□□□□□

□□□□□□□□□□□□□□□□:

$$\beta_i \sim N(\mu, \sigma^2) \quad \square\square\square \quad \hat{\theta}_i \sim N(\theta_i + \beta_i, \tau_i^2)$$

□□□ $N(a, b)$ □□□□ a □□□ b □□□□□□□□□□□□□□□□:

$$L(\mu, \sigma|\theta, \tau) \propto \prod_{i=1}^{n} \int p(\hat{\theta}_i|\beta_i, \theta_i, \hat{\tau}_i) p(\beta_i|\mu, \sigma) \mathrm{d}\beta_i$$

□□□□□□□□□□□ $\hat{\mu}$ □ $\hat{\sigma}$ □□□□□□□□□□□□□□□□□□p□□□□□□□□□□□□□□□□□

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: □□□□□□□□□□□□□□

```
  removePeopleWithPriorOutcomes = TRUE,
  washoutPeriod = 365,
  riskWindowStart = 1,
  riskWindowEnd = 0,
  endAnchor = "cohort end",
  exposureOutcomePairs = eoPairs,
  effectSizes = c(1.5, 2, 4),
  cdmVersion = cdmVersion,
  workFolder = file.path(outputFolder, "pcSynthesis"))
```

□□□□□□

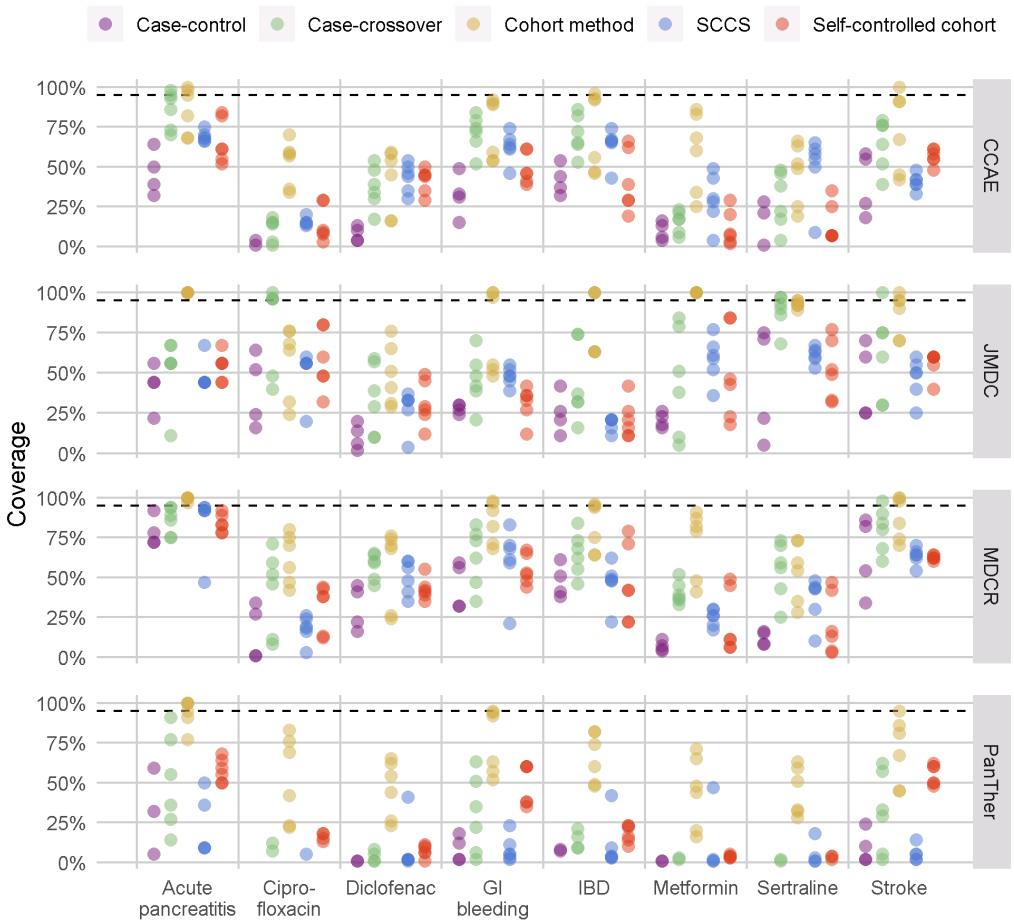


Figure 18.6: □□□□□□□□□□□□□□□□□□□□□95%□□□□□□□□□□□□□□□□□□□

Part III

OHDSI

Chapter 19

— —

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

□□□□□OHDSI□□□□□□□□

□□□□□

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

Chapter 20

– – OHDSI

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

OHDSI□□□□

20.2.2 OHDSI

□□□□□□□□□□□□□□□□□□□□□□□□ OHDSI□□□□□□□□□□□□□□□CDM□□□□□□□□□□

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

□□□□□□□□□□OHDSI□□□□□□□□□□□□□□□□□□□□□□□□□□OHDSI□□□OHDSI□□□□□□□□□□
CDM□□□OHDSI□□□□□□□□□□□□□□□□□□□

□□□□□□□□□

□□□□□□□□

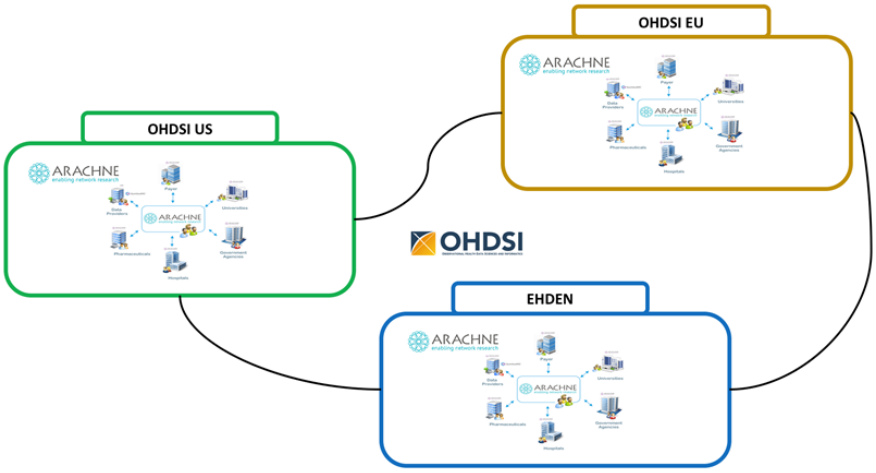


Figure 20.2: ARACHNE□□□□□□□□□□□□□□□□□□□

20.4

- OHDSI□□□□□□□□□□□□□□CDM□□□□□□□□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

```
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□□□□□□□□□□□□□□□

```
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□□□□□□□□□

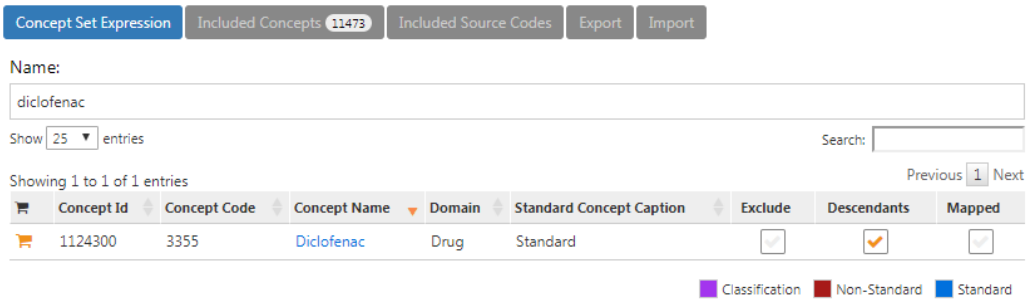


Figure E.2: □□□□□□□□□□□□□□□□□□

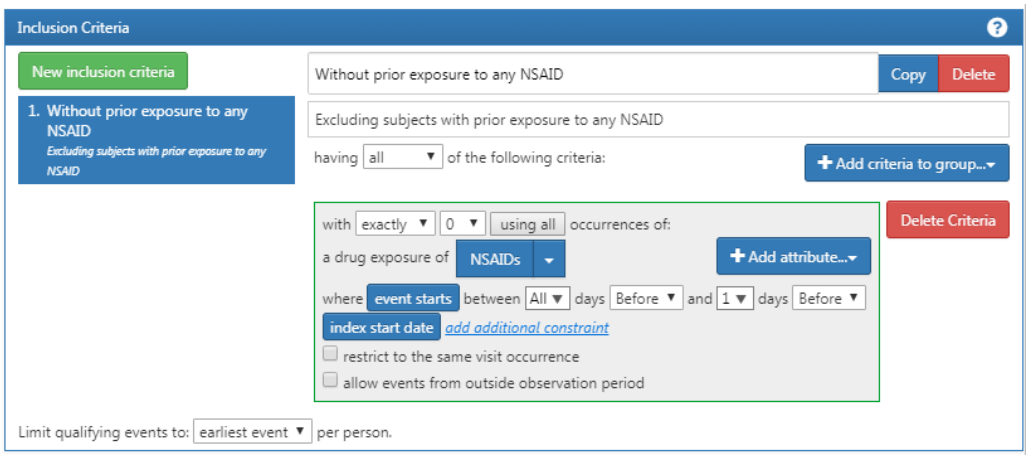


Figure E.3: NSAID□□□□□□□□□□□□□□□□

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")
```

□□□COHORT_DEFINITION_ID□□□□□□□1□

```
sql <- "INSERT INTO @cdm.cohort (
  subject_id,
  cohort_start_date,
  cohort_definition_id
  )
SELECT subject_id,
  cohort_start_date,
  CAST (1 AS INT) AS cohort_definition_id
FROM #diagnoses
INNER JOIN @cdm.visit_occurrence
  ON subject_id = person_id
    AND cohort_start_date >= visit_start_date
    AND cohort_start_date <= visit_end_date
WHERE visit_concept_id IN (9201, 9203, 262); --        ;"

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

□□□□□□□□□□□□□□□□□□□

E.7

□□ 11.1

ATLAS□ Data Sources□□□□□

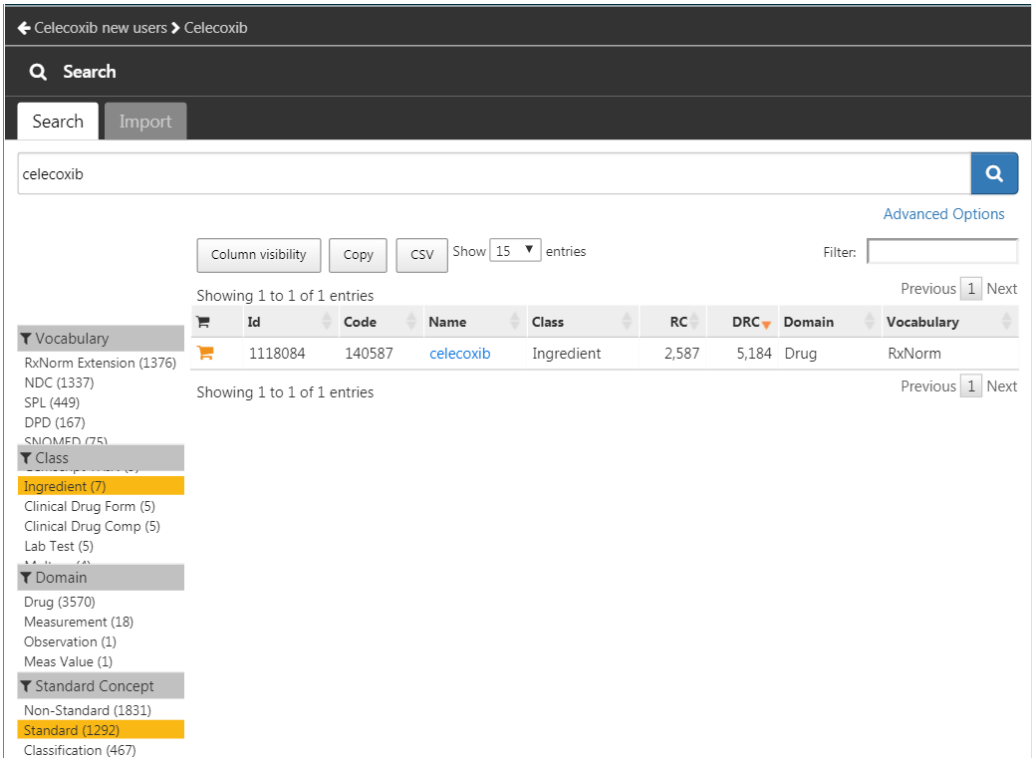


Figure E.9: □□□Celecoxib (□□□□□□)□□□□□□□□□□□□□□

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: □□□□□□

Characterization #69

Celecoxib new users characterization

Design | Executions | Utilities

Executions › Reports for SYNPUF 5%

Date: 08/23/2019 12:53 PM Design: -1840810470 Results: 2 reports

Filter panel

Cohorts: Celecoxib new users | Analyses: Condition Group Era Any Time P | Domains: Condition, Drug

CONDITION / Condition Group Era Any Time Prior

Export | Show 10 entries | Search:

Covariate	Explore	Concept ID	Count	Pct
Pain	Explore	4329041	1,140	78.62%
Pain finding at anatomical site	Explore	4132926	1,135	78.28%
Inflammation of specific body systems	Explore	4178818	1,135	78.28%
Arthropathy	Explore	73553	1,122	77.38%

Figure E.12: □□□□□□□

GI bleed › GI bleed

Search

Search | Import

Gastrointestinal hemorrhage

Advanced Options

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

Showing 1 to 15 of 25 entries | Previous 1 2 Next

Vocabulary
SNOMED (17)
ICD10CM (2)
ICD9CM (2)
DRG (2)
NDFRT (1)

Class
Clinical Finding (17)

Id	Code	Name	Class	RC	DRC	Domain	Vocabulary
192671	74474003	Gastrointestinal hemorrhage	Clinical Finding	919	37,144	Condition	SNOMED
4338544	87763006	Lower gastrointestinal hemorrhage	Clinical Finding	0	15,617	Condition	SNOMED
4100660	27719009	Acute gastrointestinal hemorrhage	Clinical Finding	0	9,852	Condition	SNOMED

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

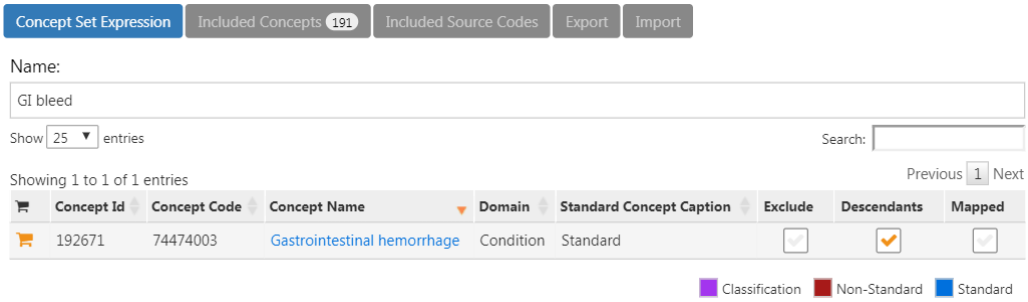


Figure E.14: □Gastrointestinal hemorrhage (□□□□□)□□□□□□□□□□□□□

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

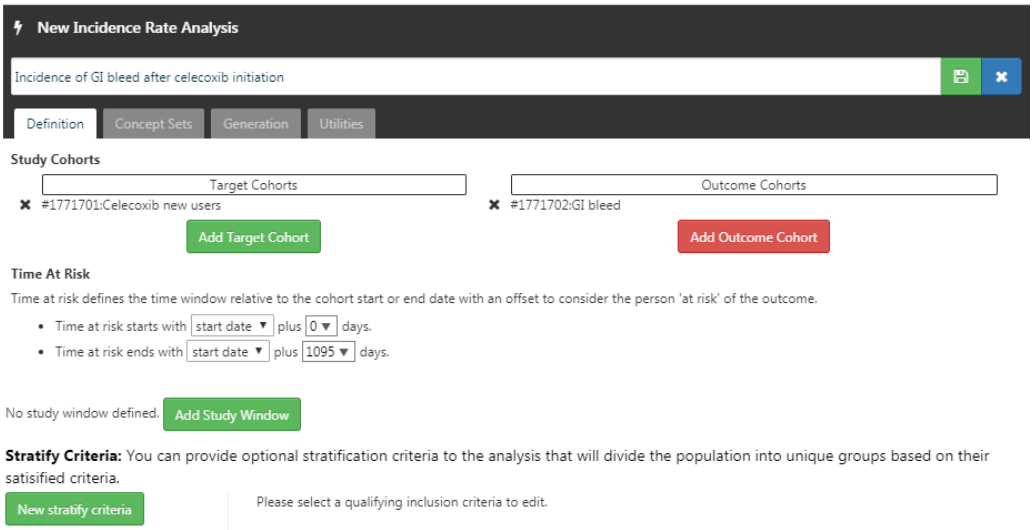


Figure E.16: □□□□□□

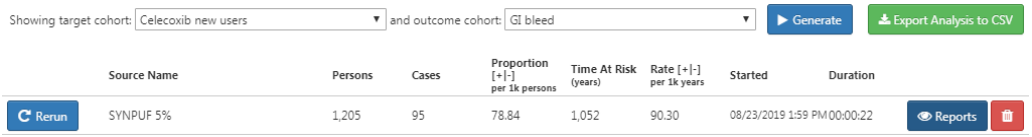


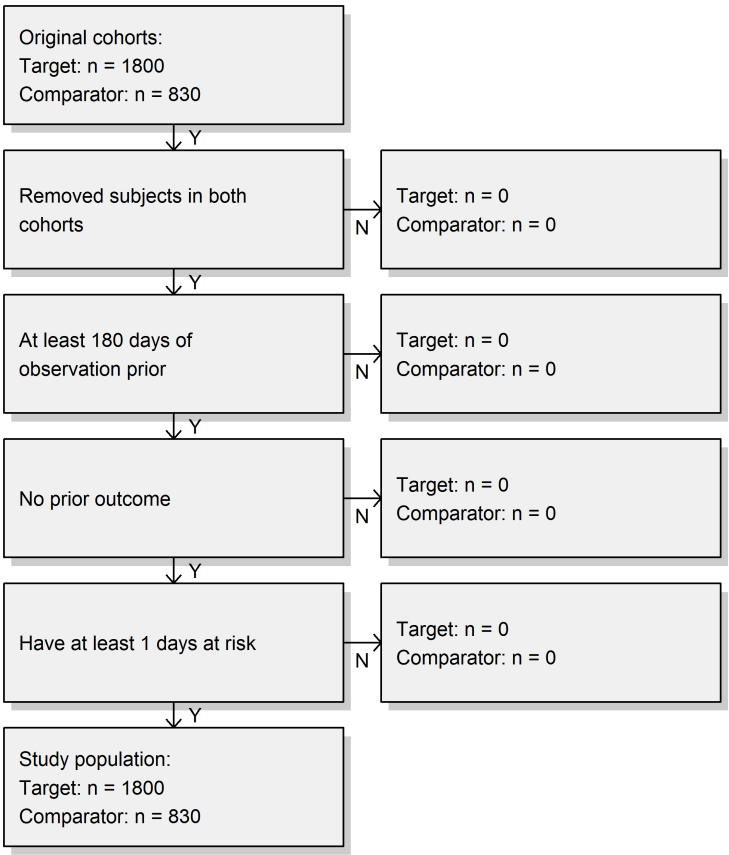
Figure E.17: □□□□□□

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.

Arnold, B. F., Ercumen, A., Benjamin-Chung, J., and Colford, J. M. (2016). Brief Report: Negative Controls to Detect Selection Bias and Measurement Bias in Epidemiologic Studies. *Epidemiology*, 27(5):637–641.

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.

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. 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.

Hill, A. B. (1965). THE ENVIRONMENT AND DISEASE: ASSOCIATION OR CAUSATION? *Proc. R. Soc. Med.*, 58:295–300.

Hripcsak, G. and Albers, D. J. (2017). High-fidelity phenotyping: richness and freedom from bias. *J Am Med Inform Assoc*.

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.

Huser, V., DeFalco, F. J., Schuemie, M., Ryan, P. B., Shang, N., Velez, M., Park, R. W., Boyce, R. D., Duke, J., Khare, R., Utidjian, L., and Bailey, C. (2016). Multisite Evaluation of a Data Quality Tool for Patient-Level Clinical Data Sets. *EGEMS (Washington, DC)*, 4(1):1239.

Huser, V., Kahn, M. G., Brown, J. S., and Gouripeddi, R. (2018). Methods for examining data quality in healthcare integrated data repositories. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 23:628–633.

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.

Lipsitch, M., Tchetgen Tchetgen, E., and Cohen, T. (2010). Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*, 21(3):383–388.

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.

Martin, R. C. (2008). *Clean Code: A Handbook of Agile Software Craftsmanship*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1 edition.

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.

Noren, G. N., Caster, O., Juhlin, K., and Lindquist, M. (2014). Zoo or savannah? Choice of training ground for evidence-based pharmacovigilance. *Drug Saf*, 37(9):655–659.

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.

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.

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.

Rubbo, B., Fitzpatrick, N. K., Denaxas, S., Daskalopoulou, M., Yu, N., Patel, R. S., Hemingway, H., Danesh, J., Allen, N., Atkinson, M., Blaveri, E., Brannan, R., Brayne, C., Brophy, S., Chaturvedi, N., Collins, R., deLusignan, S., Denaxas, S., Desai, P., Eastwood, S., Gallacher, J., Hemingway, H., Hotopf, M., Landray, M., Lyons, R., O'Neil, T., Pringle, M., Sprosen, T., Strachan, D., Sudlow, C., Sullivan, F., Zhang, Q., and Flaig, R. (2015). Use of electronic health records to ascertain, validate and phenotype acute myocardial infarction: A systematic review and recommendations. *Int. J. Cardiol.*, 187:705–711.

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.

Schuemie, M. J., Hripcsak, G., Ryan, P. B., Madigan, D., and Suchard, M. A. (2016). Ro-

bust 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., Boyce, R. D., Ryan, P. B., van der Lei, J., Rijnbeek, P. R., and Schuemie, M. J. (2016). Accuracy of an Automated Knowledge Base for Identifying Drug Adverse Reactions. *J Biomed Inform*.

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*.

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].

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.

Zaadstra, B. M., Chorus, A. M., van Buuren, S., Kalsbeek, H., and van Noort, J. M. (2008). Selective association of multiple sclerosis with infectious mononucleosis. *Mult. Scler.*, 14(3):307–313.

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□□□, 159
ACHILLES, 186
adaboost, 155
agnostic SQL, *see* SqlRender
analysis implementation, 63
APHRODITE, 95
arachne, 226
ATHENA, 33
ATLAS, 64, 160
    characterization features, 120
    □□□□□□, 66
    □□□□□□□□□, 65
    □□□□□□, 65
    □□□□□□□□, 65
    □□□, 65
    □□□□□□, 65
    □□□□□□, 64
    □□□□□□, 66
    □□□□□□□, 65
    □□□□□□, 65
    □□□□□□□□, 64
    □□□□□□□, 65
    □□□□□, 65
    □□□, 65
    □□, 65
    □□□□□□□, 65
attrition diagram, 146
AUC, 158

back-propagation, 156
baseline time, 111
best practice for network research, 226
bigknn, 155
Bill of Mortality, 31
CDM , *see* Common Data Model
characterization, 57, 111
    cohort, 111
    database level, 111
    treatment pathways, 112
classification concept, 35
Clem McDonald, 182
cohort
    entry event, 94
    exit criteria, 94
    inclusion criteria, 94
    probabilistic design, 95
    rule-based design, 94
Common Data Model, 19
    □□□□□□□□, 21
    □□□□□□, 21
    □□□□□□□, 19
    □□□□□, 19
    □□□□□□□, 21
    □□□□, 21
    □□□□□, 21
    □□□□□□, 21
    □□□□□, 19
    □□□□, 19
community, 5, 182
comparative effect estimation, 133
comparative effectiveness, *see* comparative effect estimation
concept, 33
    class, 34
    code, 35
    hierarchy, 38
    identifier, 33
    mapping, 37

relationship, 37
concept set, 95
control hypotheses, 66
convolutional neural network, 156
counterfactual, 133
covariate balance
example, 146
cross-validation, 154
Cyclops, 153

data profiling, *see* White Rabbit
data quality, 185
DatabaseConnector, 71
creating a connection, 79
querying, 79
decision boundary, 153
decision tree, 155
deep learning, 156
descriptive statistics, *see* characterization
design considerations for network research, 224
direct effect estimation, 133
disease natural history, *see* characterization
domain
concept, 34
drug utilization, 112

ETL, *see* □□□□□□□□□□□ETL□
implementations, 51
quality control, 51
□□□□□, 187
ETL design, *see* Rabbit-In-A-Hat
evidence quality, 181

FAIR, 14
feature analyses, 116
FeatureExtraction, 120
forum, 7

gradient boosting, 154

hyper-parameter, 154

incidence, 112
proportion, 113
rate, 113
index date, 111

k-nearest neighbors, 155
Kaplan-Meier plot, 146

LASSO, 153
logistic regression, 153
logistics of network research, 224

methods library, 66
minimum detectable relative risk (MDRR), 146
mission, 4

naive bayes, 155
network study, 223
neural network, 156
no free lunch, 153

objectives, 5
OHDSI Methods Benchmark, 214
OHDSI SQL, *see* SqlRender

p-value calibration, 211
Pallas system, 31
patient-level prediction, 59
perceptron, 156
person-time, 113
phenotype library, 96
PheValuator, 199
population-level estimation, 58, 133
post-index time, 111
power, 146
propensity model
example, 146
python, 154, 155

quality improvement, *see* characterization
Query Library, 71
QueryLibrary, 84

R, 71
installation, 68
Rabbit-In-A-Hat, 47
random forest, 154
recurrent neural networks, 156
regularization, 153

reliable evidence, 182
ROC, 158
running network research, 225

safety surveillance, 133
sklearn, 154
source code mapping, *see* Usagi
SQL, 71
SQL Query Library, *see* Query Library
SqlRender, 71
    debugging, 78
    parameterization, 72
    supported functions, 74
    translation, 73
standard concept, 35
Standard SQL Dialect, *see* SqlRender
standardized vocabularies, 31
    download, 33
    search, 33
structured query language, *see* SQL
study-a-thon, 13
supervised learning(□□□□), 152
survival plot, *see* Kaplan-Meier plot
system requirements, 68

tools deployment, 68
    Amazon AWS, 68
    Broadsea, 68
treatment utilization, *see* characterization
TRIPOD, 151

Usagi, 49

variance, 154
vignette, 68
vision, 4
vocabulary, 34

White Rabbit, 43

xgboost, 154
xSens cohort, 201
xSpec cohort, 200

□□□□□□□□□□, 151
    SCCS □□□□, 137
    □□□□□□□□□, 133
    □□□□□□□□□, 136
    □□□□□□□□□, 136
    □□□□□□□□□□□□□, 135
□□□□□□□□□□, 152
□□□□□□□□, 159
□□□□□□□□, 158
□□□□□□□□□□, 13
    open standards, 14
    □□□□□□□□, 14
    □□□□□□□□□□□, 14
    □□□□□□□□, 14
□□□□, 135
□□□□□□□□□, 158
□□□, 152
□□□□□□, 134
□□□□□□□□□□□□□□, *see* □□□□□□
□□□□, 93
□□□□□□□□, 133
□□□□□□, 93
□□□□□□
    □□□□□□□□□□, 8
□□□□□□□□□, 205
□□□□□
    □□, 39
□□□□□□, 93
□□□□□□□□□□, 205
□□□□□□□□□□□□□, 205
□□□□□□, *see* □□□□
□□□□□□□□□, 198
□□□□□□□□□, 151
    SCCS □□□□, 137
    □□□□□□□□□, 133
    □□□□□□□□□, 136
    □□□□□□□□□, 136
    □□□□□□□□□□□□□, 135
□□□□□
    □□□□□□□□□□, 188
    □□□□, 186
    □□□□□□□□□□, 186
    □□□□□□□□, 186
    □□□, 186
    □□□, 186
    □□, 186
    □□, 186

□□□□□□□□□□□□, 186
□□□□□□□□, *see* □□□□
□□□□□□□□□□□, 210
□□□□□□□□□□
□□□□□□□□, 136
□□□□, *see* □□□□□□□
□□□□□□□□□, 157
□□□□□□, 93
□□□□□□□□□□□□□□□□□, 205
□□□□□, 219
□□□□□□, 134
□□□□□□□□□□□, 210
□□, 210
□□□, 152
□□□□□□□, 134
□□□□□, 151
□□□□□□□□□□□□□□, *see* Common Data Model
□□□□□□□□□, 134

□□□□□□□, 151
□□□□□, 151
□□□□□□□□, 156
□□□□, 157
□□□□, 134
□□□□□□, 8

□□□□□, 135
□□□, 157
□□□, 157
□□□□□, 134
□□□□□, 134
□□, 134
□□□□, 134
□□□□□□□□, 134
□□□, 152
□□□□□□□, 135
□□□□□□□□
□□□□□□□, 23
□□, 21
□□□□□□□□, 134
□□□, 152
□□□□□□□□□□□□, 211
□□□□□, 211
□□□□□□□, 134
□□□□□□□, 151
□□, 157
□□□□□□, 209
□□
□□□□, 156
□□□□, 156
□□□□□, 157
□□□□□, 157
□□□□, 151
□□□□□, 60, 153
□□□□□□□□, 133

□□□, 157
□□□□, 43
□□-□□-□□□□□□□□□□, 136
□□□□□□□□, 136
□□□□□□□□, 136
□□□, 157
□□□, 157
□□□□□□□□, 223
□□□□□□□, 220
□□□□, 209
□□, 157

□□□□, 135
□□□□□□, 151
□□□□□□, 197
□□□□□□□□□□□□, 135
□□□□□□□□□□ (SCCS) □□□□, 137
□□□□□□□, 60
□□□□□□□, 151
□□□, 158

□□□□, 134
□□□-□□□□□□□□□, 158
□□□□□, 157