
様々なOHDSIツール

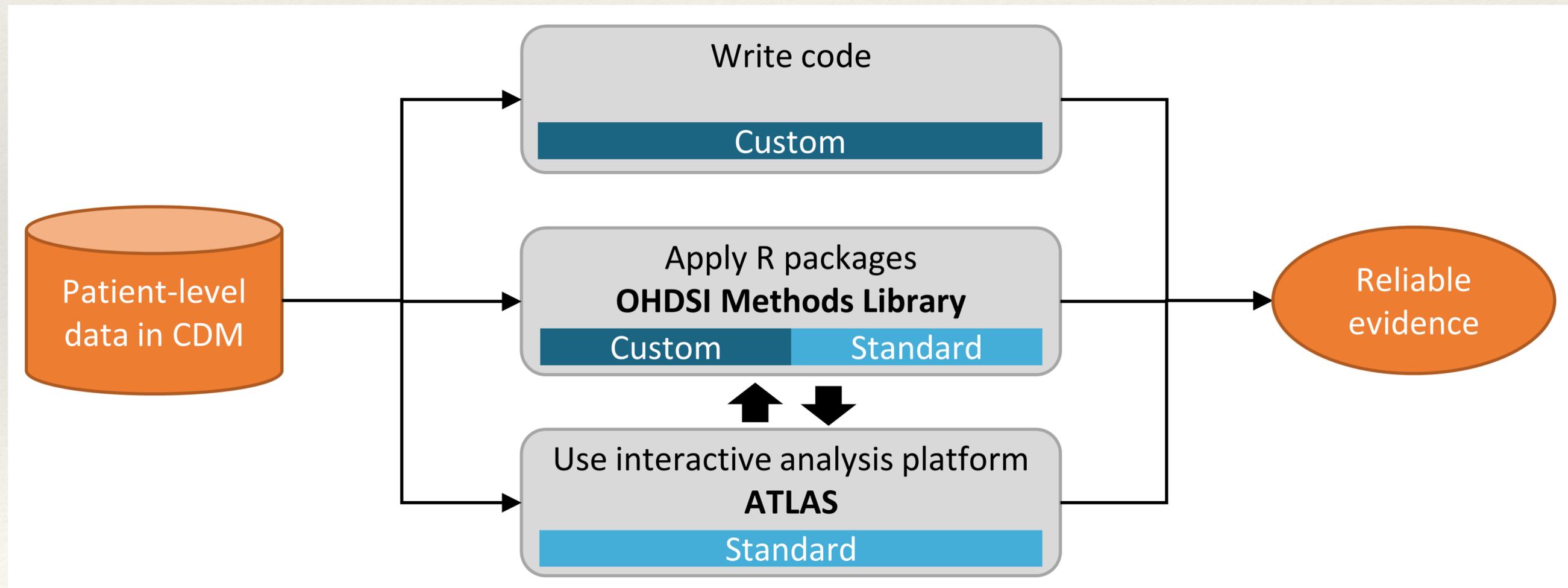
2020.11.25
OHDSI Japan Monthly meeting
岡田 昌史

本日の内容

- ❖ OHDSIツールとは
- ❖ HADES (Health Analytics Data-to-Evidence Suite)
- ❖ ATLAS
- ❖ 実際に使ってみるには

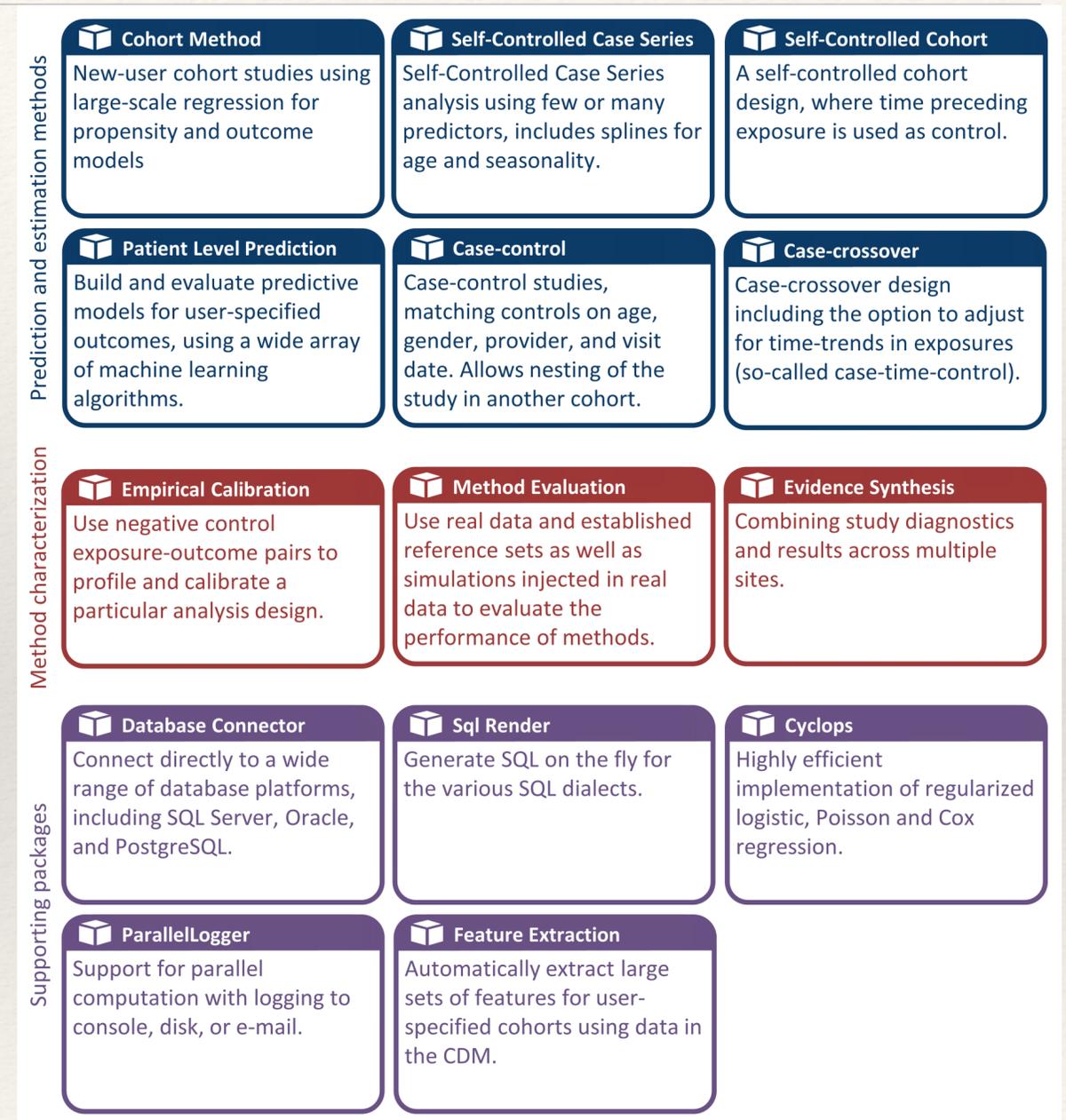
OHDSIツールとは

- ❖ OMOP CDMに準拠したデータを用いて実際のデータベース研究をするときに、よく利用される処理をRのパッケージやWebインターフェースで利用できるツールとしてまとめたもの。
- ❖ 利用しなくても研究は実施できるが、活用すればより簡単に実施できる



HADES(OHDSI Methods Library)

- ❖ CDMデータを解析するためのRパッケージ群
- ❖ データベースサーバに置かれる非常に大きいデータも、基本のR文法とSQLを知っていれば解析できる
- ❖ データベースへのアクセス方法の違いもライブラリで吸収する
- ❖ Population/Exposure/Outcomeの定義はこの関数を呼び出すRスクリプトとして実施できる



HADESを使った研究の進め方

❖ CohortMethodパッケージを使った薬剤疫学研究

パッケージCohortMethodは、2種の曝露集団間でアウトカムを比較する、という典型的なコホート研究をデータベース上で実施する際によく使われる処理をパッケージにしている

1. コホートを定義

- ❖ まず曝露群、対照群、アウトカムを定義する – “COHORT” テーブル (CDMv6) を作る
 - ❖ 抽出結果をデータベース上に置くことが必要なので、SQLで直接定義されることが多い

2. Study Populationを定義

- ❖ 群を定義したのち、共変量一覧、wash-out期間を考慮した最初の曝露の定義、指定した期間内のアウトカムの発生などをcreateStudyPopulation関数で自動的に抽出できる

3. Attrition tableの出力

4. Table 1の出力

5. Propensity Score計算

6. モデルあてはめ

7. カプランマイヤーなど

コホートの定義

データベース全体から条件に合致するコホートを定義する

曝露群のSQLによる定義 – セレコキシブ投与

```
INSERT INTO @resultsDatabaseSchema.coxibVsNonselVsGiBleed (
  cohort_definition_id,
  cohort_start_date,
  cohort_end_date,
  subject_id
)
SELECT 1, -- Exposure
  drug_era_start_date,
  drug_era_end_date,
  person_id
FROM @cdmDatabaseSchema.drug_era
WHERE drug_concept_id = 1118084;-- celecoxib
```

アウトカムのSQLによる定義 – 消化管出血による入院もしくは緊急入院

```
INSERT INTO @resultsDatabaseSchema.coxibVsNonselVsGiBleed (
  cohort_definition_id,
  cohort_start_date,
  cohort_end_date,
  subject_id
)
SELECT 3, -- Outcome
  condition_start_date,
  condition_end_date,
  condition_occurrence.person_id
FROM @cdmDatabaseSchema.condition_occurrence
INNER JOIN @cdmDatabaseSchema.visit_occurrence
  ON condition_occurrence.visit_occurrence_id = visit_occurrence.visit_occurrence_id
WHERE condition_concept_id IN (
  SELECT descendant_concept_id
  FROM @cdmDatabaseSchema.concept_ancestor
  WHERE ancestor_concept_id = 192671 -- GI - Gastrointestinal haemorrhage
)
AND visit_occurrence.visit_concept_id IN (9201, 9203);
```

コード、出力結果は CohortMethodパッケージ (Apache 2.0ライセンス) Vignette

“Single studies using the CohortMethod package” より一部改変して引用

Study Populationの定義

- ❖ 曝露(NSAIDS投与)以外のすべての状態をCovariateとして扱う。
- ❖ COHORTテーブルとして定義したtarget(曝露), comparator (対照), outcome(アウトカム) を指定
- ❖ 最初の曝露情報だけを使用(firstExposureOnly)
- ❖ 組み入れ前のwash out period は180日を指定
- ❖ targetとcomparatorの両方にいる対象者は削除(removeDuplicateSubjects)
- ❖ このほか、createStudyPopulation関数により、
 - ❖ 曝露の最小期間(minDaysAtRisk)
 - ❖ 曝露期間前後のリスク期間の幅(riskWindowStart, riskWindowEnd)
 - ❖ 曝露期間前にアウトカムがある対象者を削除(removeSubjectsWithPriorOutcome)
など指定が可能

```
nsaids <- 21603933
covSettings <- createDefaultCovariateSettings(
  excludedCovariateConceptIds = nsaids,
  addDescendantsToExclude = TRUE)

cohortMethodData <- getDbCohortMethodData(
  connectionDetails = connectionDetails,
  cdmDatabaseSchema = cdmDatabaseSchema,
  targetId = 1,
  comparatorId = 2,
  outcomeIds = 3,
  studyStartDate = "",
  studyEndDate = "",
  exposureDatabaseSchema = resultsDatabaseSchema,
  exposureTable = "cohorttbl",
  outcomeDatabaseSchema = resultsDatabaseSchema,
  outcomeTable = "cohorttbl",
  cdmVersion = cdmVersion,
  firstExposureOnly = TRUE,
  removeDuplicateSubjects = TRUE,
  restrictToCommonPeriod = FALSE,
  washoutPeriod = 180,
  covariateSettings = covSettings)
```

コード、出力結果は CohortMethodパッケージ (Apache 2.0ライセンス) Vignette

“Single studies using the CohortMethod package” より一部改変して引用

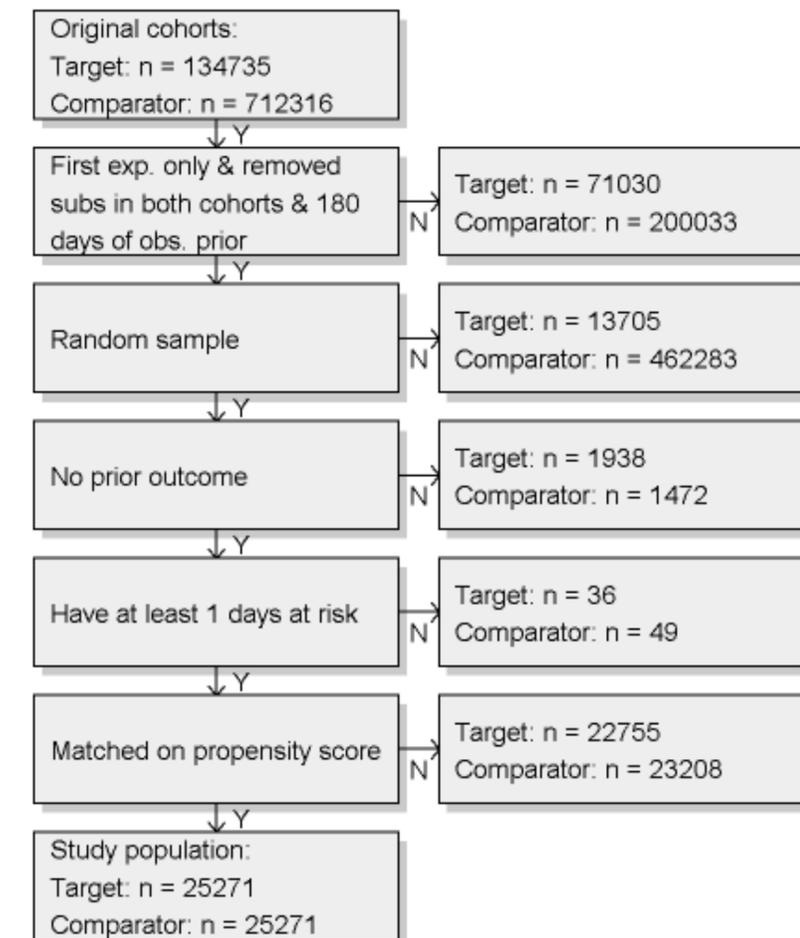
Attrition tableの出力

❖ `getAttritionTable`関数, `drawAttritionDiagram`関数で、Study population定義やマッチングによる症例数フローチャートが自動的に作成される

```
## # A tibble: 5 x 5
```

## description	targetPersons	comparatorPersons	targetExposures	comparatorExposures
## <chr>	<dbl>	<dbl>	<dbl>	<dbl>
## 1 Original cohorts	134735	712316	238173	1193013
## 2 First exp. only & removed s ...	63705	512283	63705	512283
## 3 Random sample	50000	50000	50000	50000
## 4 No prior outcome	48062	48528	48062	48528
## 5 Have at least 1 days at ris ...	48026	48479	48026	48479

`drawAttritionDiagram(matchedPop)`



コード、出力結果は CohortMethodパッケージ (Apache 2.0ライセンス) Vignette

“Single studies using the CohortMethod package” より一部改変して引用

Table 1の出力

*createCmTable1*関数により、*Patient Characteristics*のテーブルを自動出力

Characteristic	Before matching			After matching		
	Target %	Comparator %	Std. diff	Target %	Comparator %	Std. diff
Medical history: General						
Acute respiratory disease	34.3	38.6	-0.09	35.6	35.1	0.01
Attention deficit hyperactivity disorder	2.8	4.2	-0.08	3.5	3.4	0.01
Chronic liver disease	4.5	3.6	0.04	4.0	4.0	0.00
Chronic obstructive lung disease	17.3	10.4	0.20	13.3	13.7	-0.01
Crohn's disease	0.7	0.4	0.04	0.5	0.4	0.01
Dementia	2.1	1.1	0.08	1.4	1.4	0.00
Depressive disorder	33.3	29.5	0.08	31.7	32.2	-0.01
Diabetes mellitus	23.7	17.3	0.16	20.5	20.9	-0.01
Gastroesophageal reflux disease	25.1	19.3	0.14	22.3	22.7	-0.01
Gastrointestinal hemorrhage	4.2	3.1	0.06	2.0	2.1	-0.01
Human immunodeficiency virus infection	0.6	0.7	-0.01	0.6	0.7	-0.01
Hyperlipidemia	33.5	24.0	0.21	29.4	30.5	-0.02
Hypertensive disorder	48.4	37.0	0.23	42.9	44.1	-0.02
Lesion of liver	1.6	1.2	0.04	1.1	1.3	-0.01
Obesity	18.4	18.1	0.01	18.9	18.8	0.00
Osteoarthritis	46.7	30.4	0.34	39.6	41.1	-0.03
Pneumonia	5.8	3.9	0.09	4.4	4.4	0.00

コード、出力結果は CohortMethodパッケージ (Apache 2.0ライセンス) Vignette

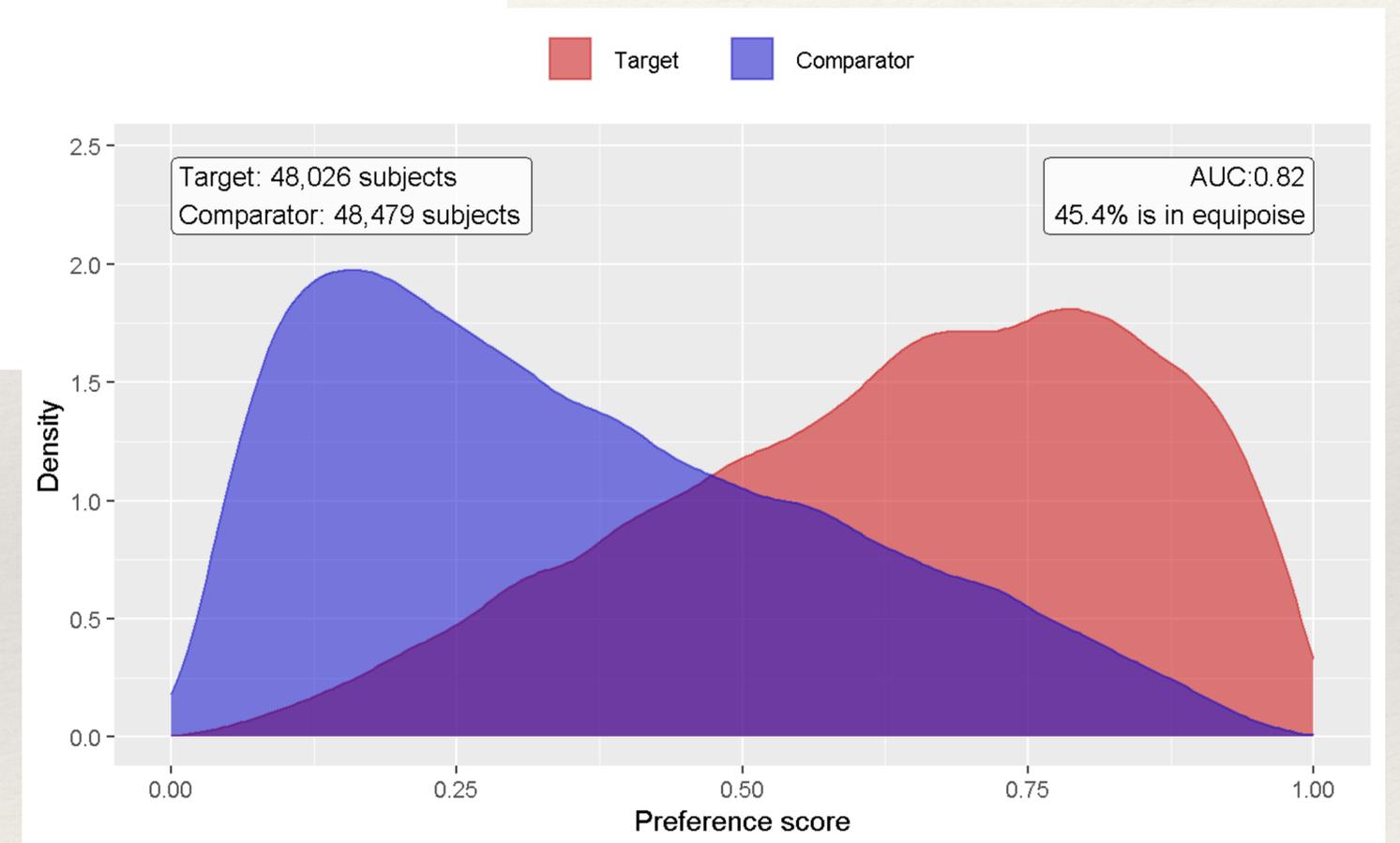
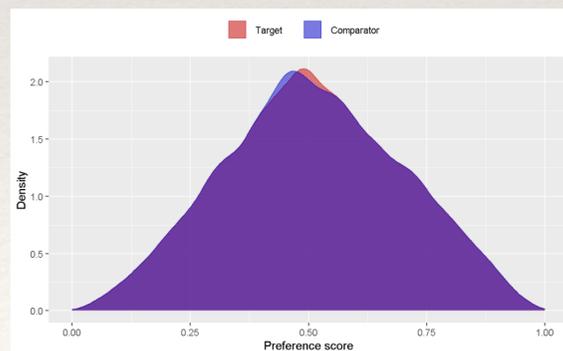
“Single studies using the CohortMethod package” より一部改変して引用

Propensity Score Matching

多数のCovariatesに対してPropensity scoreを算出可能

```
ps <- createPs(cohortMethodData = cohortMethodData, population = studyPop)
plotPs(ps,
  scale = "preference",
  showCountsLabel = TRUE,
  showAucLabel = TRUE,
  showEquiposeLabel = TRUE)
```

matchOnPs関数を使うことで、
Propensity Score Matchingも可能



コード、出力結果は CohortMethodパッケージ (Apache 2.0ライセンス) Vignette
“Single studies using the CohortMethod package” より一部改変して引用

ATLAS



- ❖ データベーススタディの基礎となる、コホート定義からデータマネジメント、解析まで、OHDSIのツールへの統合されたWebインターフェース
- ❖ OMOP CDM準拠データであればプログラミング不要で対話的にデータ処理が可能
- ❖ 作成したコホートに対する記述統計や特性の視覚化、一般的なデータの妥当性検査、基本的な解析、処理のRスクリプトでのエクスポートなどの機能をもつ

Cohort #1771427

EXAMPLE: new users of ACE inhibitors as first-line mono-therapy for hypertension

Definition ? Concept Sets Generation Reporting Export

enter a cohort definition description here

Cohort Entry Events ?

Events having any of the following criteria:

with continuous observation of at least days before and days after event index

Limit initial events to: per person.

Restrict initial events

+ Add Initial Event ▾

Add Condition Era
Find patients with specific diagnosis era.

Add Condition Occurrence
Find patients with specific diagnoses.

Add Death
Find patients based on death.

Add Device Exposure
Find patients based on device exposure.

Add Dose Era
Find patients with dose eras.

Add Drug Era
Find patients with exposure to drugs over time.

Add Drug Exposure
Find patients with exposure to specific drugs or drug classes.

Inclusion Criteria

New inclusion criteria

Limit qualifying events to: per person.

Cohort Exit

ATLASでのコホート定義

- ❖ まず、定義に用いるConceptのセットを用意する
- ❖ コホートに入るイベント、続いて絞り込むInclusion CriteriaをConcept setを用いて定義する
- ❖ Demographics, Condition, Drug Exposureなどから条件を定義可能

Inclusion Criteria

New inclusion criteria

- has hypertension diagnosis in 1 yr prior to treatment
- Has no prior antihypertensive drug exposures in medical history
- Is only taking ACE as monotherapy, with no concomitant combination treatments**

Is only taking ACE as monotherapy, with no concomitant combination treatments **Copy** **Delete**

enter an inclusion rule description

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

with **exactly** **1** using distinct occurrences of: **+ Add attribute...**

a drug era of **Hypertension drugs**

where **event starts** between **0** days **Before** and **7** days **After**

index start date [add additional constraint](#)

allow events from outside observation period

Delete Criteria

Limit qualifying events to: **earliest event** per person.

← EXAMPLE: new users of ACE inhibitors as first-line mono-therapy for hypertension → ACE Inhibitors

Search **Import**

Search **Import**

ace inhibitors **Q**

Advanced Options

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

Showing 1 to 9 of 9 entries **Previous** **1** **Next**

	Id	Code	Name	Class	RC	DRC	Domain	Vocabulary
	21601784	C09AA	ACE inhibitors, plain	ATC 4th	0	507,772	Drug	ATC
(1)	21601783	C09A	ACE INHIBITORS, PLAIN	ATC 3rd	0	507,772	Drug	ATC
s (1)	21601802	C09BA	ACE inhibitors and diuretics	ATC 4th	0	10,982	Drug	ATC
1)								
	21601801	C09B	ACE INHIBITORS, COMBINATIONS	ATC 3rd	0	10,982	Drug	ATC
(4)								

- Cohort generationは時間がかかるので、定義と実行は別のインターフェースとなっている
- 定義をJSONやSQLで書き出し、共有することが可能
- 異なるデータベースに対して同一定義を適用してコホートを作ることができる

Cohort Characterization

❖ 変数ごとの記述統計(Table 1相当)

CONDITION / Condition Occurrence Long Term / 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	Count	Pct	
Tachycardia	Explore	444070	17,322	1.04%	9,000	1.04%	
Cardiomegaly	Explore	314658	20,958	1.26%	8,000	1.26%	
Cardiac arrhythmia	Explore	44784217	30,474	1.83%	13,000	1.83%	

Showing 1 to 3 of 3 entries (filtered from 206 total entries)

DEMOGRAPHICS / Demographics Age

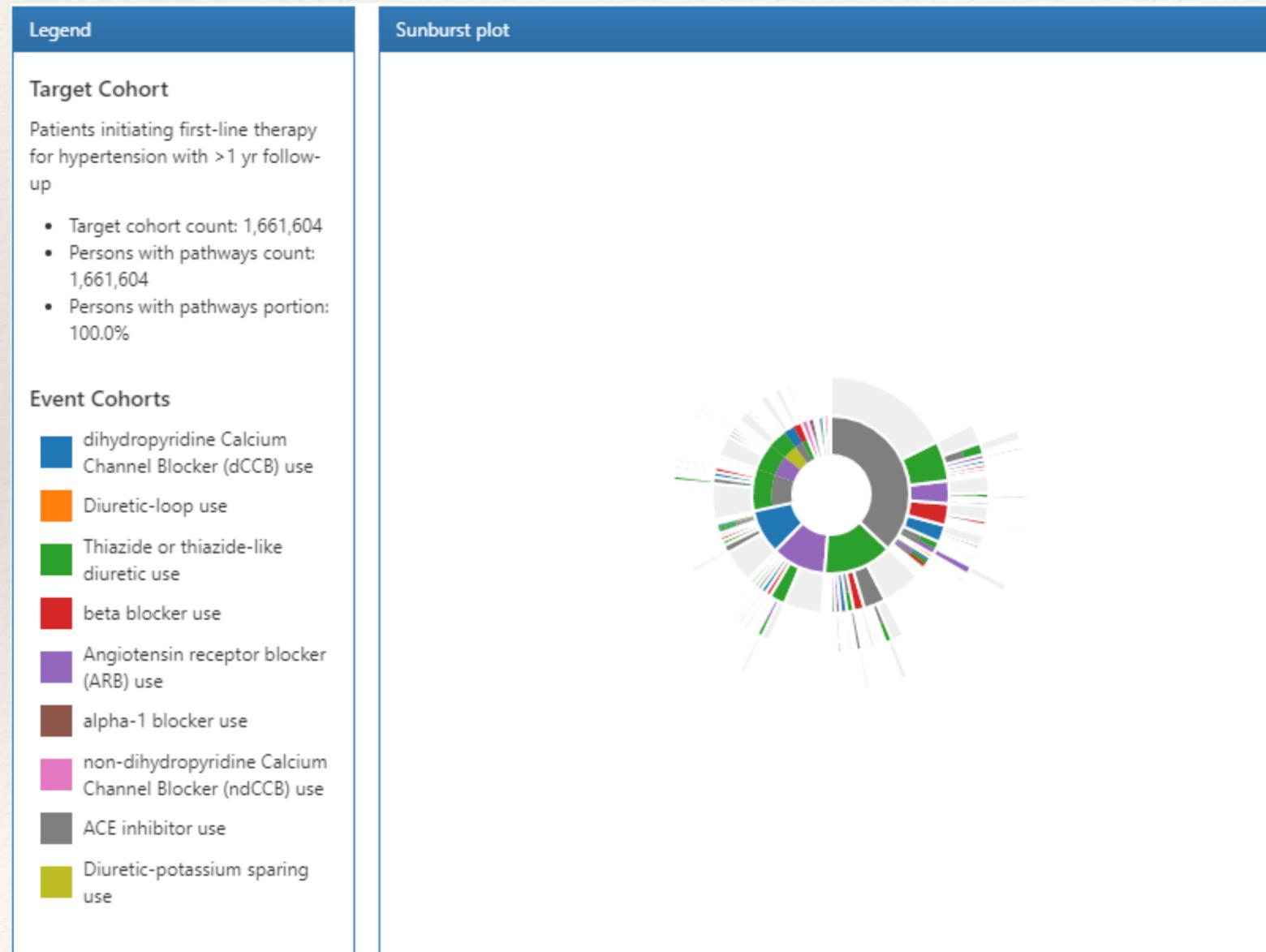
Export Export comparison Show 10 entries Search:

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

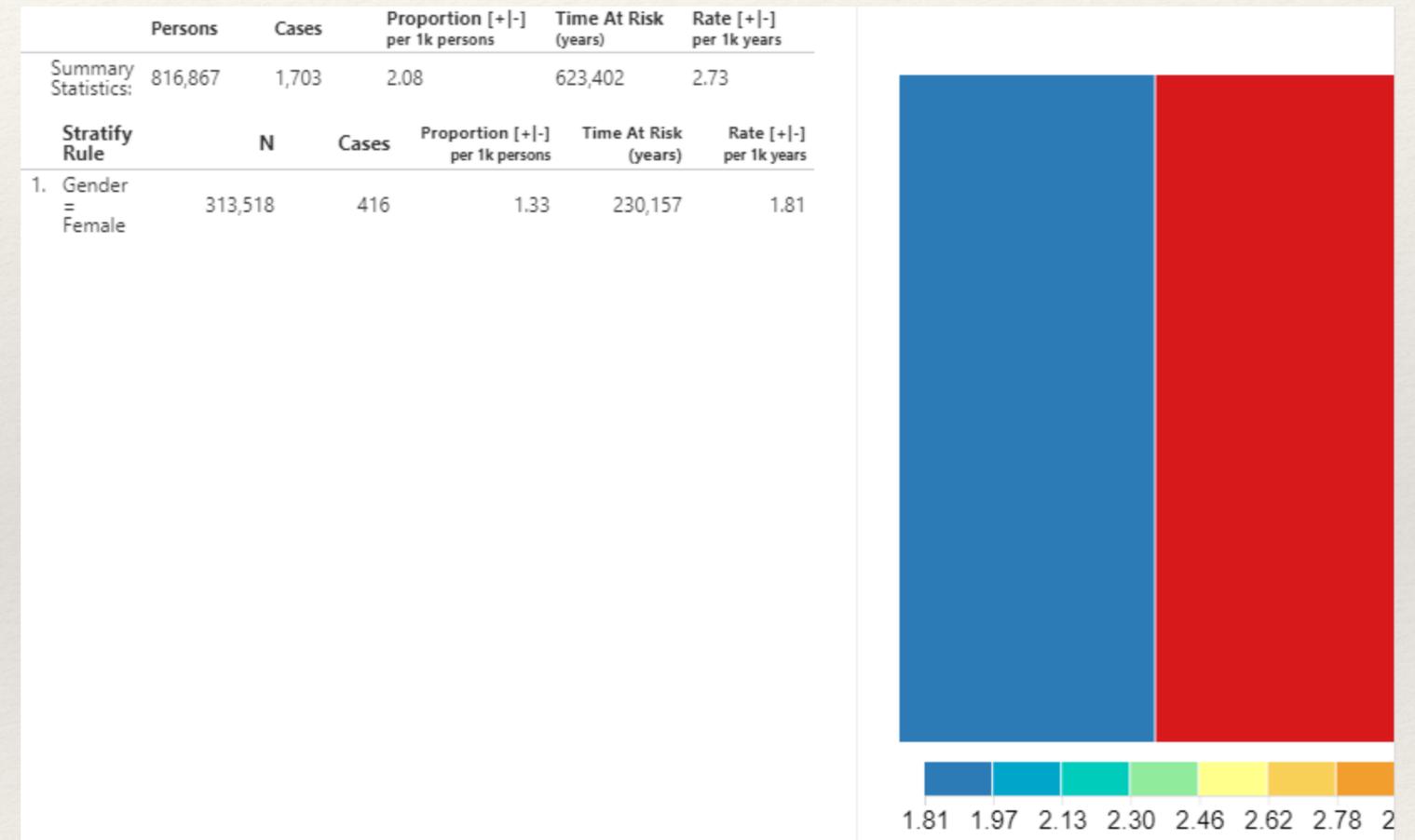
Showing 1 to 2 of 2 entries Previous 1 Next

Cohort Pathways, Incidence Analysis

- ❖ Target cohort に対して、Event cohortを複数定義することで、イベントが発生した順序ごとに割合を可視化する



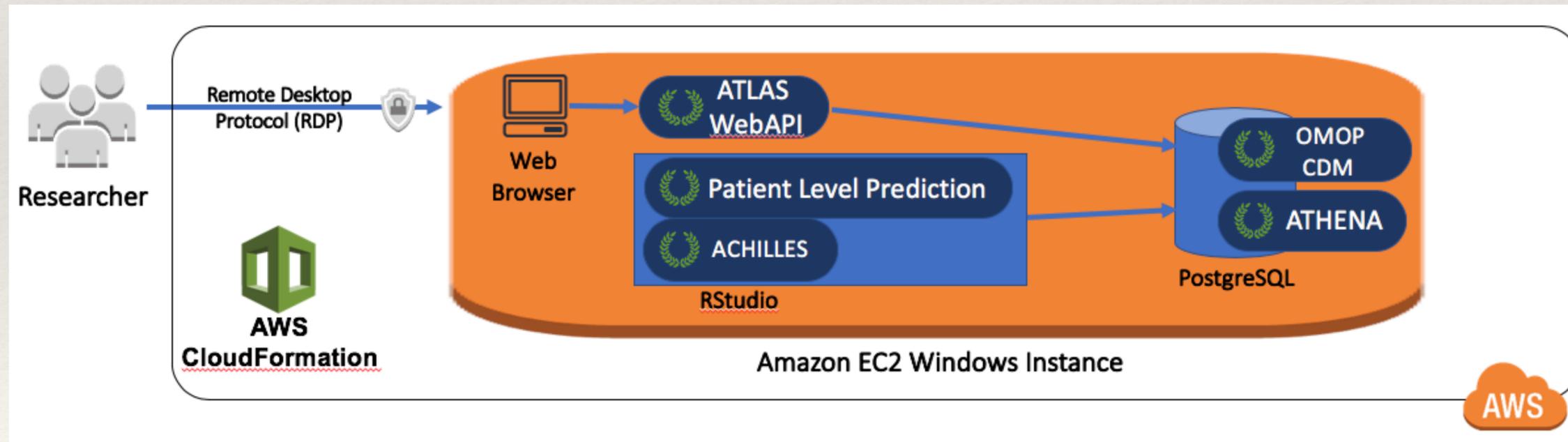
- 層化変数を指定して、イベント発生率を比較して可視化する



画面イメージは The Book of OHDSI (CC0-1.0 ライセンス) より引用

実際に使ってみるには

- ❖ OHDSIツールを利用するには、CDMに従ったデータベースをデータベース管理ソフトウェアに導入する必要がある
- ❖ データ自体は仮想データセット作成ツール(Synthea), 匿名化公開データ(desynpuf) などから得られるが、環境設定はそれなりに手間がかかる
- ❖ OHDSI In a Box <https://github.com/OHDSI/OHDSI-in-a-Box> を用いることで、AWS上にすでに環境がセットアップされた仮想マシンを構築することができる。Rstudio, synthea と desynpufデータが格納されたPostgreSQL, ATLAS がセットアップされており、無料枠でも十分利用できる



まとめ

- ❖ OHDSI CDM形式で作成されたデータベースに対して、HADES, ATLASなどのツールを用いることで、データベース研究を効率的に実施できる
- ❖ HADESはRのパッケージとして、ATLASはWebツールとして利用される。
- ❖ Target cohort, Comparator cohort, Outcome cohortの3つのコホートを定義した上で、Study populationを定義する方法(CohortMethod)を用いることで、よく使われる分析を少ないコードで実行することができる