

Curso OMOP

Phenotyping: Herramientas OMOP para definir problemas de salud II: CohortDiagnostics

Introducción

- Hemos elaborado unas definiciones para nuestras cohortes de interés (exposición/*outcomes*/criterios de exclusión...) a partir de Atlas, CodeListGenerator, desarrollando el código en R
- ¿Podríamos mejorar nuestras definiciones?
¿Hemos olvidado de algo? o hemos introducido algún error?

→ Podemos hacer una exploración inicial de como “funcionan” nuestras definiciones en nuestra base de datos con CohortDiagnostics

— Qué es CohortDiagnostics?

- Paquete de R para evaluar algoritmos de fenotipos desarrollados para bases de datos en OMOP CDM
- Permite identificar posibles errores en los conceptos de inclusión, cuantificar (counts), ver si hay “overlap” entre distintas cohortes, etc.
- Los resultados se pueden explorar a través de una ShinyApp
- Se pueden combinar resultados de más de una base de datos
- Se considera una buena práctica usarlo cuando hacemos un estudio en red

Ventajas

- “Fácil” de implementar en R (problemas más comunes son de instalación del paquete)
- Los resultados los puede visualizar una persona sin conocimientos de OMOP/R (por ej, clínicos)
- Facilita la identificación de errores
- Facilita la comparación de definiciones entre cohortes y entre bases de datos → muy importante cuando hacemos estudios en red
- Nos permite mejorar nuestras definiciones (proceso iterativo)

<https://ohdsi.github.io/CohortDiagnostics/>

CohortDiagnostics

 R-CMD-check passing  codecov 89%

CohortDiagnostics is part of [HADES](#).

Introduction

CohortDiagnostics is an R utility package for the development and evaluation of phenotype algorithms for OMOP CDM compliant data sets. This package provides a standard, end to end, set of analytics for understanding patient capture including data generation and result exploration through an R Shiny interface. Analytics computed include cohort characteristics, record counts, index event misclassification, captured observation windows and basic incidence proportions for age, gender and calendar year. Through the identification of errors, CohortDiagnostics enables the comparison of multiple candidate cohort definitions across one or more data sources, facilitating reproducible research.

Características

- Mostrar las reglas de inclusión y atrición
- Ver los códigos fuente utilizados en una definición
- Encontrar códigos “huérfanos”
- Mostrar incidencias de la cohorte (*aproximación*)
- Desglosar los eventos “índice” en los conceptos específicos que los desencadenaron
- Calcular la superposición entre dos cohortes
- Caracterizar cohortes y comparar cohortes

Un ejemplo

Cohort Diagnostics

- </> Cohort Definition
- 🗪 Concepts in Data Source
- 🏠 Orphan Concepts
- ☰ Cohort Counts
- + Incidence Rate
- 🕒 Time Distributions
- 📊 Index Event Breakdown
- Cohort Overlap
- 👤 Cohort Characterization
- 👥 Compare Characterization
- ⚙️ Meta data

Cohort Definition

Export Cohorts Zip

Search

	Cohort Id ↑	Cohort Name ↑
	<input type="text"/>	<input type="text"/>
<input type="radio"/>	1	covid19_diagnosis
<input type="radio"/>	2	covid19_diagnosis_positive_test
<input type="radio"/>	3	covid19_positive_test
<input type="radio"/>	4	Grupo1
<input type="radio"/>	5	Grupo2
<input type="radio"/>	6	Grupo3
<input type="radio"/>	7	Grupo4

1-7 of 7 rows Show 20 ▾

Previous **1** Next

Cohort Definition

Cohort Diagnostics

</> Cohort Definition

Concepts in Data Source

Orphan Concepts

Cohort Counts

+ Incidence Rate

Time Distributions

Index Event Breakdown

Cohort Overlap

Cohort Characterization

Compare Characterization

Meta data

Cohort Definition

Export Cohorts Zip

Search

	Cohort Id ↑	Cohort Name ↑
	<input type="text"/>	<input type="text"/>
<input type="radio"/>	1	covid19_diagnosis
<input type="radio"/>	2	covid19_diagnosis_positive_test
<input type="radio"/>	3	covid19_positive_test
<input type="radio"/>	4	Grupo1
<input type="radio"/>	5	Grupo2
<input type="radio"/>	6	Grupo3
<input type="radio"/>	7	Grupo4

1-7 of 7 rows Show 20

Previous 1 Next

Listado de las cohortes incluidas

Concepts in Data Source

</> Cohort Definition

Concepts in Data Source

Orphan Concepts

Cohort Counts

+ Incidence Rate

Time Distributions

Index Event Breakdown

Cohort Overlap

Cohort Characterization

Compare Characterization

Meta data

Database(s)

SIDIAP

Cohort

C1: covid19_diagnosis

Concepts in Data Source

C1: covid19_diagnosis

Source fields

Standard fields

Both

Persons

Records

Search

SIDIAP				
Concept Id ↓	Concept Name ↓	Vocabulary Id ↓	Persons (104,084) ↓	Records (115,676) ↓
37311061	COVID-19	SNOMED	99,002	750,538
439676	Coronavirus infection	SNOMED	7,596	55,660
4100065	Disease due to Coronaviridae	SNOMED	1,207	2,238
3661408	Pneumonia caused by SARS-CoV-2	SNOMED	566	742
320651	Severe acute respiratory syndrome	SNOMED	202	697
40479642	Pneumonia due to Severe acute respiratory syndrome co...	SNOMED	34	105

1–6 of 6 rows Show 20

Previous 1 Next

Download as CSV

Conceptos incluidos – con counts de Personas/registros

Orphan Concepts

Concepts in Data Source

Orphan Concepts

Cohort Counts

Incidence Rate

Time Distributions

Index Event Breakdown

Cohort Overlap

Cohort Characterization

Compare Characterization

Meta data

Database(s)

SIDIAP

Cohort

C1: covid19_diagnosis

C1: covid19_diagnosis

Filters

All

Standard Only

Non Standard Only

Display

All

Persons

Records

Sort By

Concept Id

order

ascending

descending

Download

Concept Id	Concept Name	Vocabulary Id	Concept Code	Standard Concept	SIDIAP	
					Records	Persons
37310257	Measurement of Severe acute respiratory syndrome coronavirus 2 antigen	SNOMED	1240471000000102	Standard	338,411	157,923
37003518	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML Injectable Suspension	RxNorm	2470234	Standard	175,623	135,817
37003436	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension	RxNorm	2468235	Standard	400,035	188,348
766503	Post COVID-19 condition, unspecified	ICD10CM	U09.9	Non-standard	1,054	598
739906	SARS-COV-2 (COVID-19) vaccine, vector -	RxNorm	2479835	Standard	13,407	13,384

Conceptos que podrían faltar en nuestra definición

Cohort Counts

</> Cohort Definition

🗃 Concepts in Data Source

🗑 Orphan Concepts

☰ Cohort Counts

+ Incidence Rate

🕒 Time Distributions

📊 Index Event Breakdown

○ Cohort Overlap

👤 Cohort Characterization

👥 Compare Characterization

⚙ Meta data

Database(s)

SIDIAP

Cohorts

C1: covid19_diagnosis, C2: c ▲

Cohort Counts

+

C1: covid19_diagnosis
C2: covid19_diagnosis_positive_test
C3: covid19_positive_test

Display

☒ Both ☐ Persons ☐ Records

Search

SIDIAP				
	Cohort Id	Cohort Name	Persons	Records
	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="radio"/>	1	covid19_diagnosis	104,084	115,676
<input type="radio"/>	2	covid19_diagnosis_positive_test	121,489	143,309
<input type="radio"/>	3	covid19_positive_test	98,953	112,241

1-3 of 3 rows Show 20 ▼

Previous **1** Next

Download as CSV

Incidence Rate

Concepts in Data Source

Orphan Concepts

Cohort Counts

+ Incidence Rate

Time Distributions

Index Event Breakdown

Cohort Overlap

Cohort Characterization

Compare Characterization

Meta data

Database(s)

SIDIAP

Cohorts

C1: covid19_diagnosis, C2: c ▲

Stratify by

☐ Age ☒ Sex ☒ Calendar Year

☐ Use same y-scale across databases

Filter By Sex

Female, Male

Minimum person years

1000

Minimum subject count

Plot row height (pixels)

497

Filter By Calendar Year

Year 2006

Year 2020

Year 2023

Generate Report

Selected cohorts :

C1: covid19_diagnosis

C2: covid19_diagnosis_positive_test

C3: covid19_positive_test

Selected databases :

SIDIAP

Plot

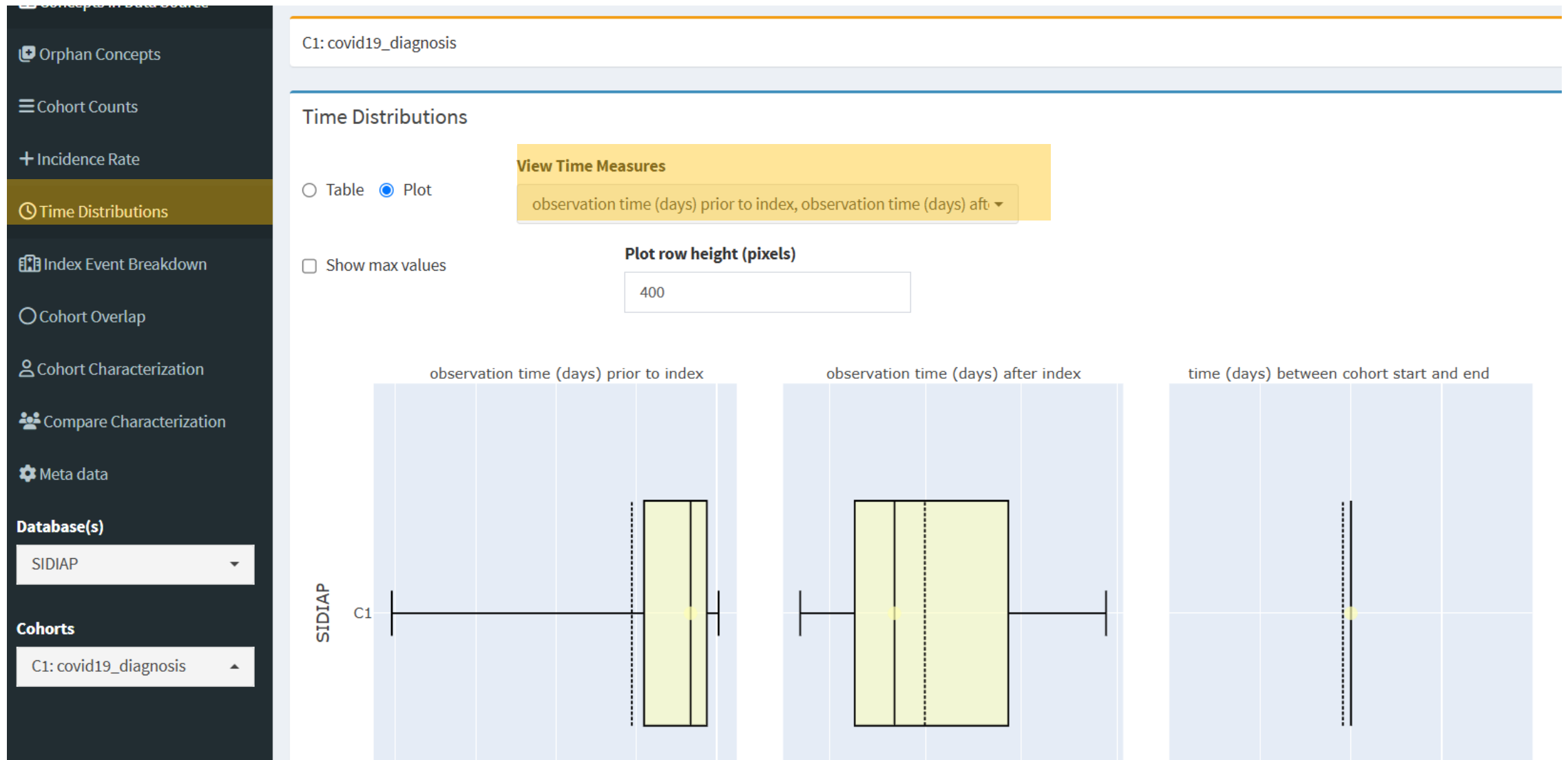
Table

Incidencia estratificada por edad, sexo, año + filtros / en tabla o gráfico

Incidence Rate



Time Distribution



Distribuciones temporales: días de observación antes/después Index date

Index Event Breakdown

Cohort Definition

Concepts in Data Source

Orphan Concepts

Cohort Counts

Incidence Rate

Time Distributions

Index Event Breakdown

Cohort Overlap

Cohort Characterization

Compare Characterization

Meta data

Database(s)

SIDIAP

Cohort

C1: covid19_diagnosis

Index Events

+

C1: covid19_diagnosis

Concept type

☐ All ☒ Standard concepts ☐ Non Standard Concepts

Display

☐ Both ☐ Records ☒ Persons ☒ Show persons as percentage

Search

					SIDIAP (104,084)
Concept Id ↓	Concept Name ↓	Concept Code ↓	Domain Field ↓	Vocabulary Id ↓	Persons ↓
37311061	COVID-19	840539006	condition_concept_id	SNOMED	95.0%
439676	Coronavirus infection	186747009	condition_concept_id	SNOMED	7.0%
4100065	Disease due to Coronaviridae	27619001	condition_concept_id	SNOMED	0.7%
3661408	Pneumonia caused by SAR...	882784691000119100	condition_concept_id	SNOMED	0.3%
320651	Severe acute respiratory sy...	398447004	condition_concept_id	SNOMED	0.2%
40479642	Pneumonia due to Severe a...	441590008	condition_concept_id	SNOMED	0.0%
40479642	Pneumonia due to Severe a...	441590008	condition_source_concept...	SNOMED	0.0%

1-7 of 7 rows

Show 20

Previous 1 Next

Download as CSV

Qué condition ha hecho que la persona entrase en la cohorte

Cohort Overlap

</> Cohort Definition

🗃 Concepts in Data Source

🗑 Orphan Concepts

☰ Cohort Counts

+ Incidence Rate

🕒 Time Distributions

📊 Index Event Breakdown

🔍 Cohort Overlap

👤 Cohort Characterization

👥 Compare Characterization

⚙ Meta data

Database(s)
SIDIAP

Cohorts
C1: covid19_diagnosis, C2: c

Cohort Overlap (subjects) +

C1: covid19_diagnosis
C2: covid19_diagnosis_positive_test

Plot

Table

☒ Show As Percentage ☐ Show Cohort Ids

Search

Target Cohort Name ↕	Comparator Cohort Name ↕	Database Name ↕	↕ T Only	↕ C Only	↕ Both	Total Subjects
▼ covid19_diagnosis (1)						
	▼ covid19_diagnosis_positive_test (1)					
		SIDIAP	0.0%	14.3%	85.7%	121,489

1-1 of 1 rows Show 20 ▼

Previous 1 Next

Grado de solapamiento entre cohortes → útil para ver qué “aporta” algún concept más

Cohort Characterization

</> Cohort Definition

Concepts in Data Source

Orphan Concepts

Cohort Counts

+ Incidence Rate

Time Distributions

Index Event Breakdown

Cohort Overlap

Cohort Characterization

Compare Characterization

Meta data

Cohort Characterization

Table type

☐ Pretty ☒ Raw

Select Cohort

covid19_diagnosis

Select Database (s)

SIDIAP

Temporal Window (s)

T (-365d to -31d), T (-30d to -1d), T (0d to 0d)

Analysis name

Chads2, Chads2Vasc, CharlsonIndex, Cohort

Domain name

Cohort, Condition, Demographics, Device, D

Covariate type(s)

☒ All ☐ Proportion ☐ Continuous

Display

☐ Mean and Standard Deviation ☒ Mean only

Subset to Concept Set

Nothing selected

Percentage displayed where only proportional data is selected

Generate Table

Cohort : covid19_diagnosis

Database(s) : SIDIAP

Características de la cohorte (descriptivo edad, sexo, diagnósticos, medicamentos...)

Compare Characterization

</> Cohort Definition

📊 Concepts in Data Source

👤 Orphan Concepts

☰ Cohort Counts

+ Incidence Rate

🕒 Time Distributions

📅 Index Event Breakdown

🔄 Cohort Overlap

👤 Cohort Characterization

👥 Compare Characterization

⚙️ Meta data

Compare Cohort Characterization +

Compare cohort characterization

Target Cohort	Target Database	Comparator Cohort	Comparator Database
covid19_diagnosis ▼	SIDIAP ▼	covid19_positive_test ▼	SIDIAP ▼

Temporal Window (s)
T (1d to 30d), T (31d to 365d) ▼

Analysis name
Chads2, Chads2Vasc, CharlsonIndex, Cohort ▼

Domain name
Condition ▼

Min Covariate Mean
0,005

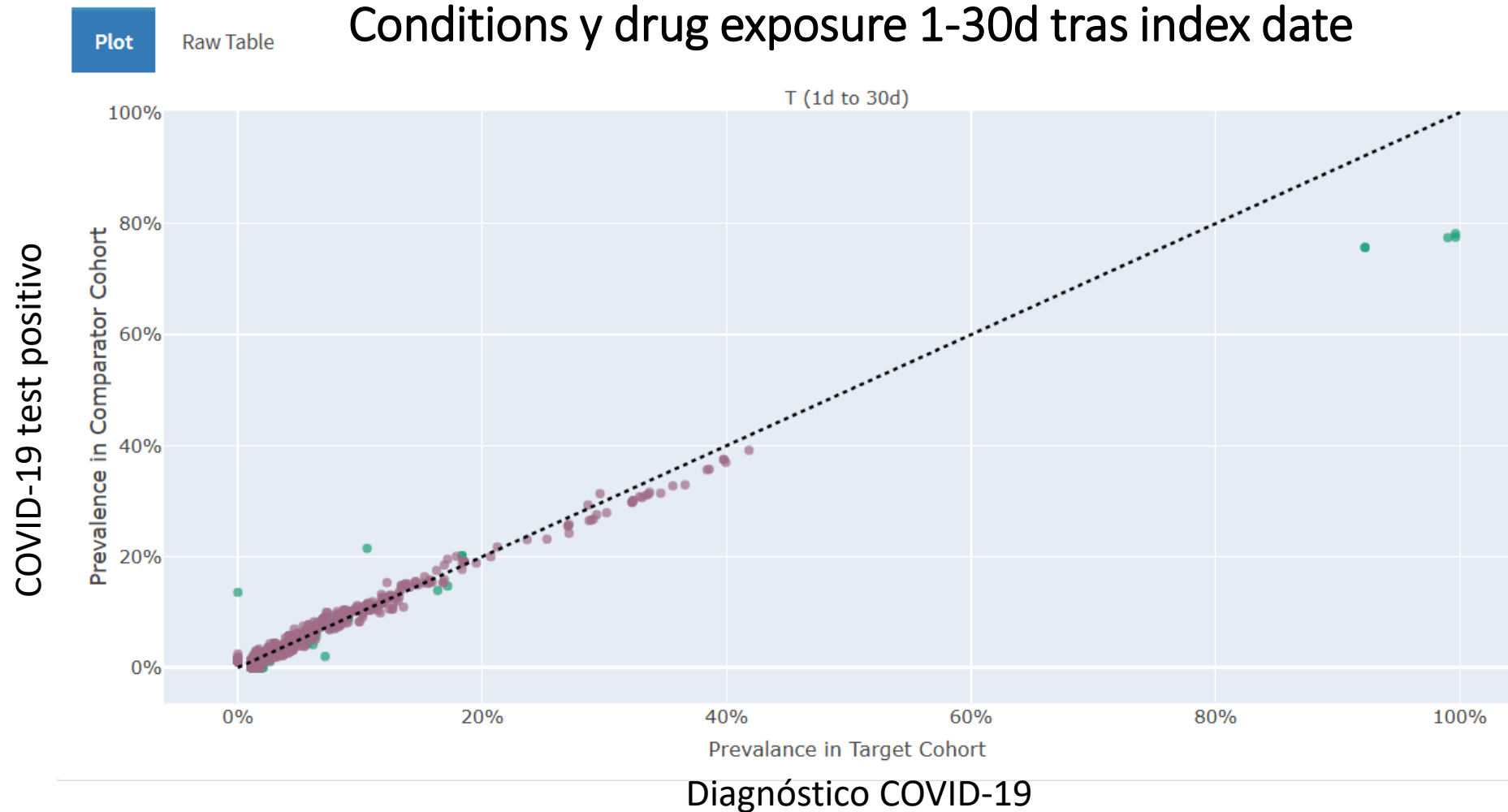
Generate Report

Target Cohort : covid19_diagnosis C1
Comparator Cohort : covid19_positive_test C3

Target Database : SIDIAP
Comparator Database : SIDIAP

Comparar características entre distintas cohortes o entre distintas bases de datos

Compare Characterization



Otro ejemplo con varias bases de datos

<https://data.ohdsi.org/LegendT2dmClassCohortExplorer/>

LEGEND-T2DM Class Cohorts

Cohort Definition

Concepts in Data Source

Orphan Concepts

Cohort Counts

Incidence Rate

Time Distributions

Inclusion Rule Statistics

Index Event Breakdown

Visit Context

Cohort Characterization

Temporal Characterization

C001: DPP4I main(101100000)

Inclusion Rule Events

☒ All ☐ Meet ☐ Gain ☐ Remain ☐ Totals

Show 100 entries

Search:

Download

Rule Sequence ID	Rule Name	OptumEHR			
		Meet	Gain	Remain	Total
All	All	All	All	All	All
0	No prior GLP-1 receptor agonist exposure	471,387	2,239	509,933	471,387
1	No prior SGLT-2 inhibitor exposure	462,429	8,088	509,933	433,535
2	No prior SU exposure	270,782	70,947	509,933	232,227
3	No prior other anti-diabetic exposure	434,881	8,948	509,933	207,902
4	Prior metformin use	279,892	77,995	509,933	97,939
5	No prior insulin use or combo initiation	360,363	20,753	509,933	77,186

Showing 1 to 6 of 6 entries

Previous1Next

Ejecución en R

1. Instalar el paquete
2. Conectarnos a la BD
3. Cargar las cohortes de referencia
4. Generar (instanciar) las cohortes
5. Ejecutar CohortDiagnostics
6. Guardar resultados
7. Ejecutar la ShinyApp

Ejecución en R

1. Instalar el paquete
2. Conectarnos a la BD
3. Cargar las cohortes de referencia
4. Generar (instanciar) las cohortes
5. Ejecutar CohortDiagnostics
6. Guardar resultados
7. Ejecutar la ShinyApp

A nivel práctico, modificaremos el código lo mínimo posible:

- ubicar los jsons en la carpeta correspondiente
- elegir qué diagnósticos ejecutar

executeDiagnostics

```
executeDiagnostics(cohortDefinitionSet,  
                  connectionDetails = connectionDetails,  
                  cohortTable = cohort_table_name,  
                  cohortDatabaseSchema = results_database_schema,  
                  cdmDatabaseSchema = cdm_database_schema,  
                  exportFolder = here("Results"),  
                  databaseId = "SIDIAP",  
                  minCellCount = 5,  
                  runInclusionStatistics = TRUE,  
                  runOrphanConcepts = TRUE,  
                  runTimeDistributions = TRUE,  
                  runVisitContext = TRUE,  
                  runBreakdownIndexEvents = TRUE,  
                  runIncidenceRate = TRUE,  
                  runTimeSeries = TRUE,  
                  runCohortOverlap = TRUE,  
                  runCohortCharacterization = TRUE,  
                  runTemporalCohortCharacterization = TRUE)
```

executeDiagnostics

```
executeDiagnostics(cohortDefinitionSet,  
                  connectionDetails = connectionDetails,  
                  cohortTable = cohort_table_name,  
                  cohortDatabaseSchema = results_database_schema,  
                  cdmDatabaseSchema = cdm_database_schema,  
                  exportFolder = here("Results"),  
                  databaseId = "SIDIAP",  
                  minCellCount = 5,  
                  runInclusionStatistics = TRUE,  
                  runOrphanConcepts = TRUE,  
                  runTimeDistributions = TRUE,  
                  runVisitContext = TRUE,  
                  runBreakdownIndexEvents = TRUE,  
                  runIncidenceRate = TRUE,  
                  runTimeSeries = TRUE,  
                  runCohortOverlap = TRUE,  
                  runCohortCharacterization = TRUE,  
                  runTemporalCohortCharacterization = TRUE)
```

Cohortes que
caracterizaremos
(resultado del paso 3)

executeDiagnostics

```
executeDiagnostics(cohortDefinitionSet,  
                  connectionDetails = connectionDetails,  
                  cohortTable = cohort_table_name,  
                  cohortDatabaseSchema = results_database_schema,  
                  cdmDatabaseSchema = cdm_database_schema,  
                  exportFolder = here("Results"),  
                  databaseId = "SIDIAP",  
                  minCellCount = 5,  
                  runInclusionStatistics = TRUE,  
                  runOrphanConcepts = TRUE,  
                  runTimeDistributions = TRUE,  
                  runVisitContext = TRUE,  
                  runBreakdownIndexEvents = TRUE,  
                  runIncidenceRate = TRUE,  
                  runTimeSeries = TRUE,  
                  runCohortOverlap = TRUE,  
                  runCohortCharacterization = TRUE,  
                  runTemporalCohortCharacterization = TRUE)
```

Tabla donde están las Cohortes instanciadas (resultado del paso 4)

executeDiagnostics

```
executeDiagnostics(cohortDefinitionSet,  
                  connectionDetails = connectionDetails,  
                  cohortTable = cohort_table_name,  
                  cohortDatabaseSchema = results_database_schema,  
                  cdmDatabaseSchema = cdm_database_schema,  
                  exportFolder = here("Results"),  
                  databaseId = "SIDIAP",  
                  minCellCount = 5,  
                  runInclusionStatistics = TRUE,  
                  runOrphanConcepts = TRUE,  
                  runTimeDistributions = TRUE,  
                  runVisitContext = TRUE,  
                  runBreakdownIndexEvents = TRUE,  
                  runIncidenceRate = TRUE,  
                  runTimeSeries = TRUE,  
                  runCohortOverlap = TRUE,  
                  runCohortCharacterization = TRUE,  
                  runTemporalCohortCharacterization = TRUE)
```

Donde guardaremos
los Resultados
(carpeta creada antes)

executeDiagnostics

```
executeDiagnostics(cohortDefinitionSet,  
                  connectionDetails = connectionDetails,  
                  cohortTable = cohort_table_name,  
                  cohortDatabaseSchema = results_database_schema,  
                  cdmDatabaseSchema = cdm_database_schema,  
                  exportFolder = here("Results"),  
                  databaseId = "SIDIAP",  
                  minCellCount = 5,  
                  runInclusionStatistics = TRUE,  
                  runOrphanConcepts = TRUE,  
                  runTimeDistributions = TRUE,  
                  runVisitContext = TRUE,  
                  runBreakdownIndexEvents = TRUE,  
                  runIncidenceRate = TRUE,  
                  runTimeSeries = TRUE,  
                  runCohortOverlap = TRUE,  
                  runCohortCharacterization = TRUE,  
                  runTemporalCohortCharacterization = TRUE)
```

Con qué nombre
quiero que
aparezca mi BD

executeDiagnostics

```
executeDiagnostics(cohortDefinitionSet,  
                  connectionDetails = connectionDetails,  
                  cohortTable = cohort_table_name,  
                  cohortDatabaseSchema = results_database_schema,  
                  cdmDatabaseSchema = cdm_database_schema,  
                  exportFolder = here("Results"),  
                  databaseId = "SIDIAP",  
                  minCellCount = 5,  
                  runInclusionStatistics = TRUE,  
                  runOrphanConcepts = TRUE,  
                  runTimeDistributions = TRUE,  
                  runVisitContext = TRUE,  
                  runBreakdownIndexEvents = TRUE,  
                  runIncidenceRate = TRUE,  
                  runTimeSeries = TRUE,  
                  runCohortOverlap = TRUE,  
                  runCohortCharacterization = TRUE,  
                  runTemporalCohortCharacterization = TRUE)
```

Seleccionar qué
Diagnósticos
quiero ejecutar