OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

# OHDSI community efforts on COVID-19 disease natural history: Status update and look forward to 'life after COVID'

Patrick Ryan, PhD

Janssen Research and Development
Columbia University Irving Medical Center

on behalf of OHDSI community:
CHARYBDIS study leads: Anthony Sena, Kristin Kostka,
Talita Duarte-Salles, Albert Prats-Uribe

# Open collaboration requires FULL transparency in every step of the research process

- Protocol and analysis source code freely available and directly downloadable:
  https://github.com/ohdsi-studies/Covid19CharacterizationCharybdis

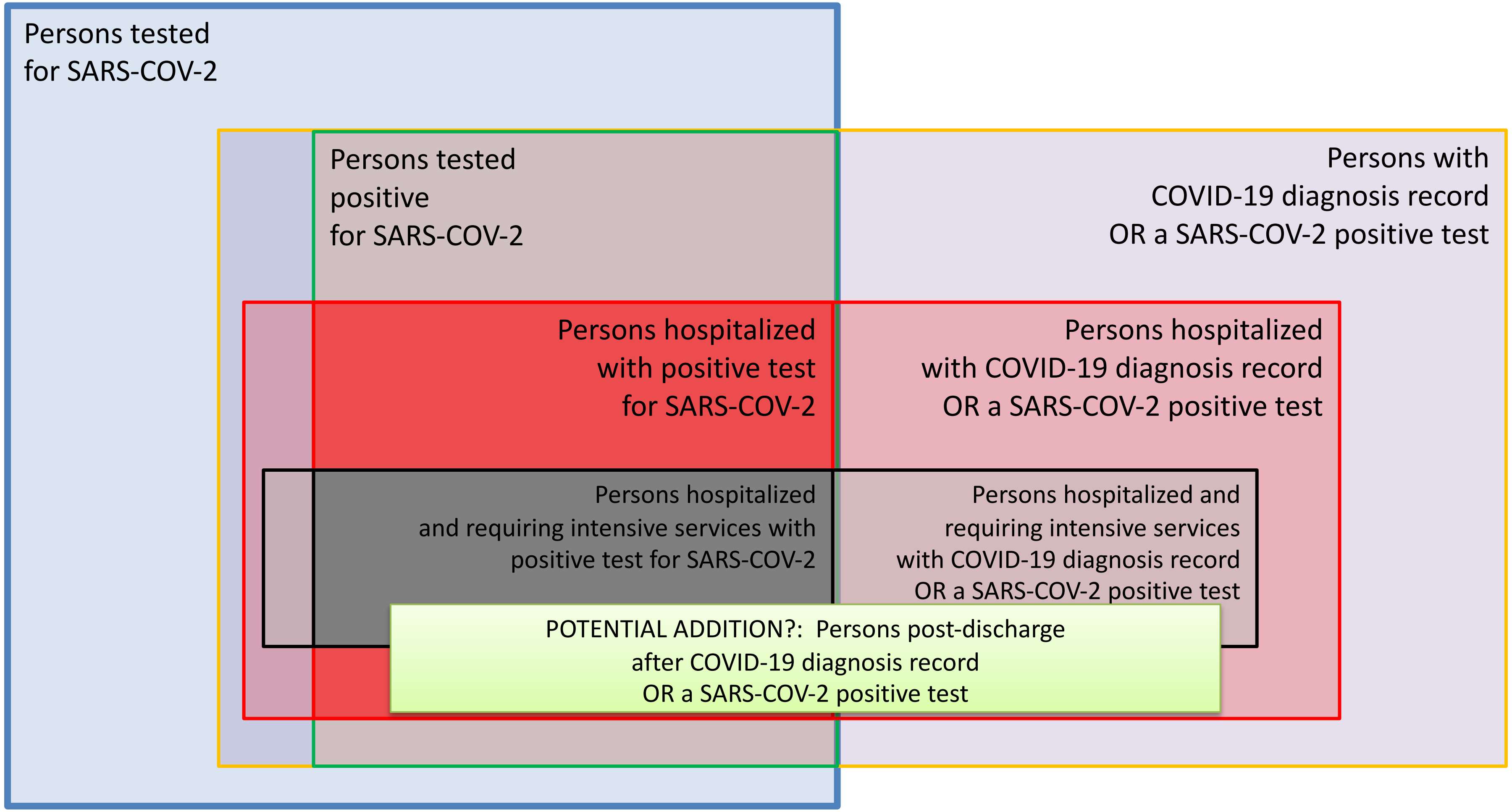
- Phenotype definitions are both human-readable and computer-executable using ATLAS against any OMOP CDM:
  https://atlas.ohdsi.org/

- All analysis results will be available for public exploration through interactive R shiny application:
  http://data.ohdsi.org/Covid19CharacterizationCHARYBDIS/

- The study is a living evidence repository: any data partners can execute analysis and share aggregate results at any point, including updates as data accumulate

**Join the Journey!**

# CHARYBDIS target cohorts



Persons tested for SARS-COV-2

Persons tested positive for SARS-COV-2

Persons with COVID-19 diagnosis record OR a SARS-COV-2 positive test

Persons hospitalized with positive test for SARS-COV-2

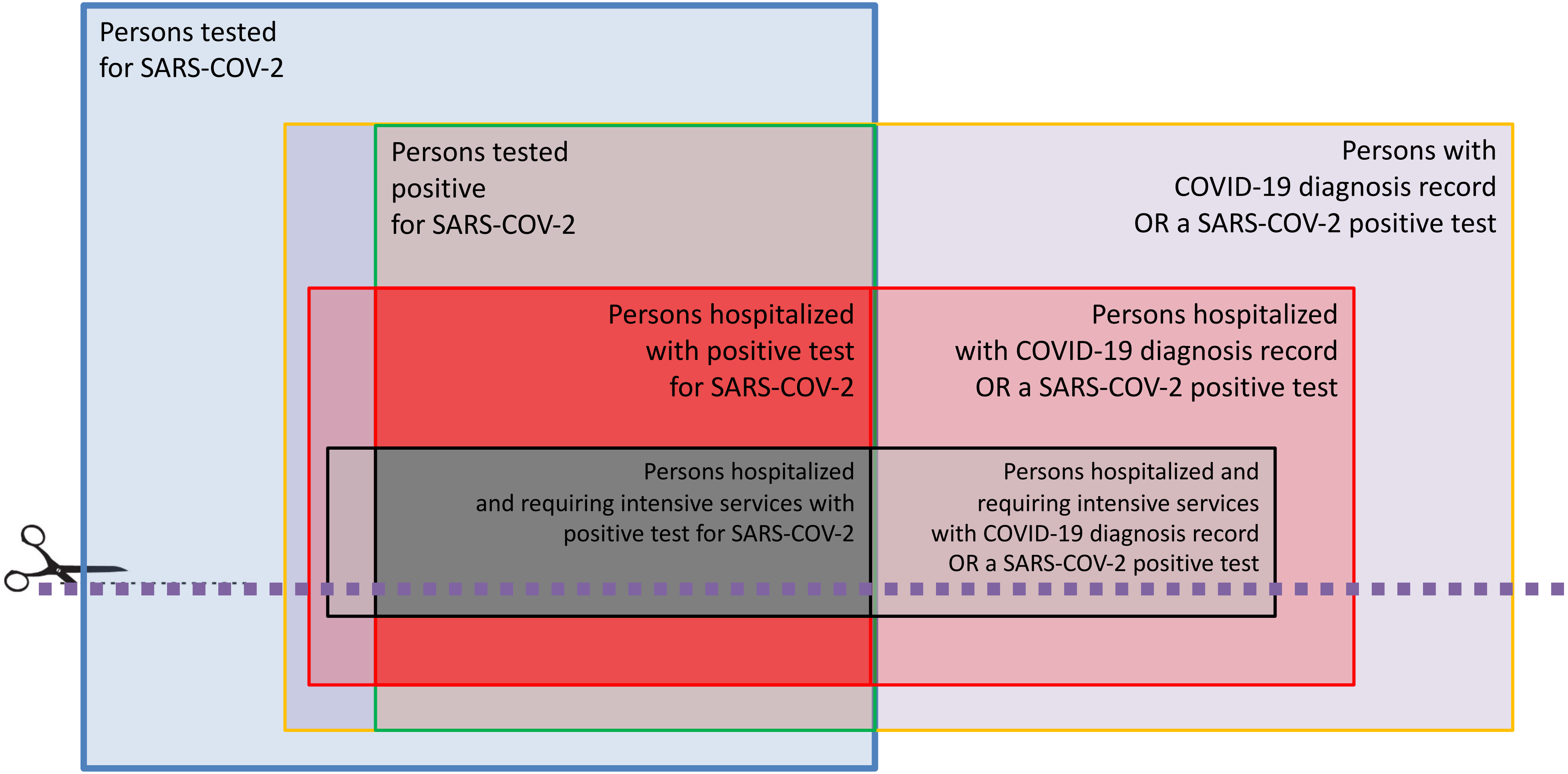
Persons hospitalized with COVID-19 diagnosis record OR a SARS-COV-2 positive test

Persons hospitalized and requiring intensive services with positive test for SARS-COV-2

Persons hospitalized and requiring intensive services with COVID-19 diagnosis record OR a SARS-COV-2 positive test

POTENTIAL ADDITION?: Persons post-discharge after COVID-19 diagnosis record OR a SARS-COV-2 positive test

# CHARYBDIS subgroup cohorts



Persons tested for SARS-COV-2

Persons tested positive for SARS-COV-2

Persons with COVID-19 diagnosis record OR a SARS-COV-2 positive test

Persons hospitalized with positive test for SARS-COV-2

Persons hospitalized with COVID-19 diagnosis record OR a SARS-COV-2 positive test

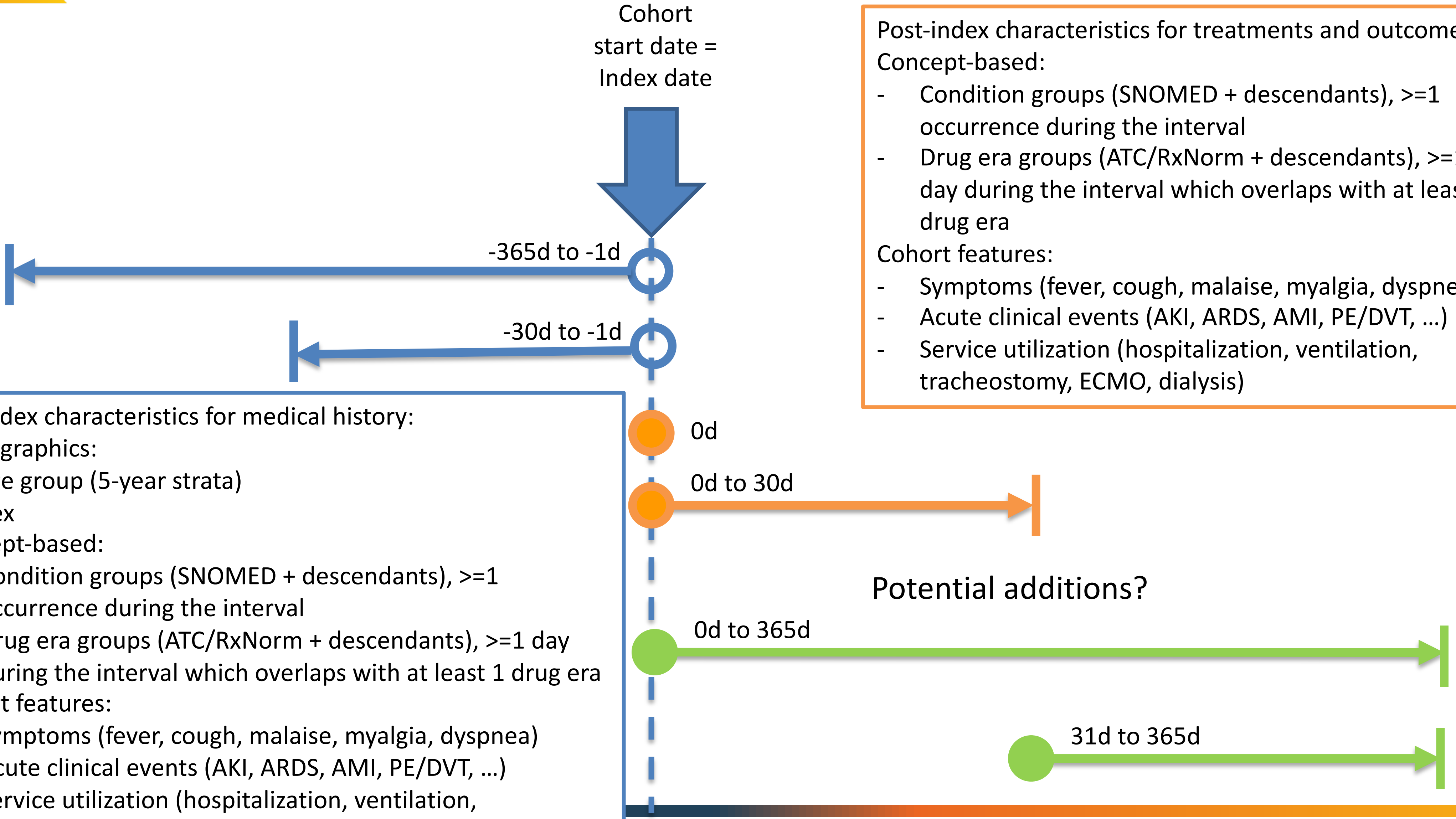
Persons hospitalized and requiring intensive services with positive test for SARS-COV-2

Persons hospitalized and requiring intensive services with COVID-19 diagnosis record OR a SARS-COV-2 positive test

Stratification cohorts:

- Age: <18, >65
- Gender: Female/Male
- Race: Black/White
- Index month
- Hypertension
- Type 2 Diabetes
- Heart disease
- Obesity
- Asthma
- COPD
- Chronic kidney disease
- End stage renal disease
- Cancer
- Autoimmune conditions
- Dementia
- HIV
- Pregnant women
- **Follow-up time: >=30d**

# CHARYBDIS Time windows



Post-index characteristics for treatments and outcomes:
Concept-based:
- Condition groups (SNOMED + descendants), >=1 occurrence during the interval
- Drug era groups (ATC/RxNorm + descendants), >=1 day during the interval which overlaps with at least 1 drug era

Cohort features:
- Symptoms (fever, cough, malaise, myalgia, dyspnea)
- Acute clinical events (AKI, ARDS, AMI, PE/DVT, …)
- Service utilization (hospitalization, ventilation, tracheostomy, ECMO, dialysis)

Pre-index characteristics for medical history:
Demographics:
- Age group (5-year strata)
- Sex

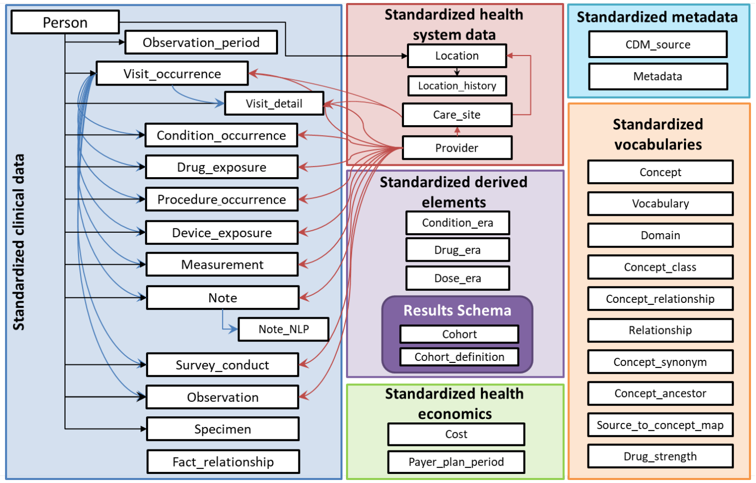
Concept-based:
- Condition groups (SNOMED + descendants), >=1 occurrence during the interval
- Drug era groups (ATC/RxNorm + descendants), >=1 day during the interval which overlaps with at least 1 drug era

Cohort features:
- Symptoms (fever, cough, malaise, myalgia, dyspnea)
- Acute clinical events (AKI, ARDS, AMI, PE/DVT, …)
- Service utilization (hospitalization, ventilation, tracheostomy, ECMO, dialysis)

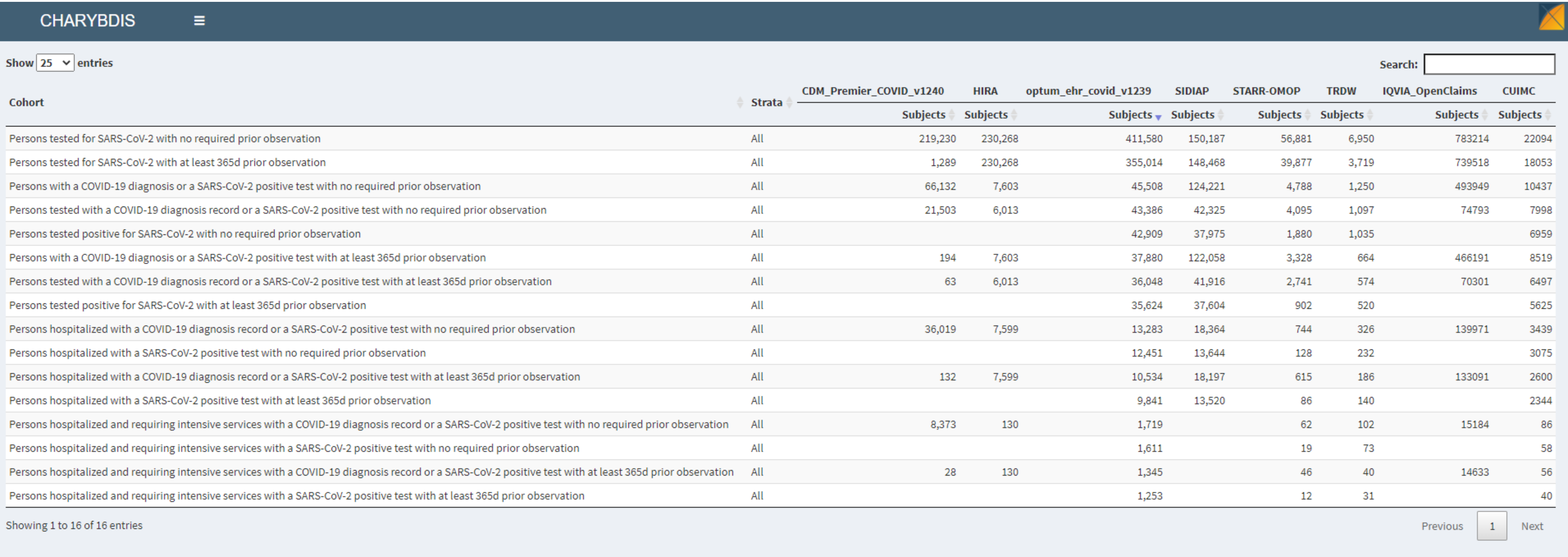
# Data partners contributing to CHARYBDIS thusfar

| Database name | Geography | Data type |
|---|---|---|
| Premier | US (National) | Hospital billing |
| Optum EHR | US (National) | Electronic health records |
| Iqvia Open Claims | US (National) | Administrative claims |
| VINCI (VA) | US (National) | Electronic health records |
| STARR (Stanford) | US (CA) | Electronic health records |
| TRDW (Tufts) | US (MA) | Electronic health records |
| CUIMC (Columbia) | US (NY) | Electronic health records |
| SIDIAP | Spain | Electronic health records |
| SIDIAP-H | Spain | EHR-hospital linkage |
| HM Hospitales | Spain | Hospital billing |
| ICPI | Netherlands | Electronic health records |
| CPRD | UK | Electronic health records |
| HIRA | South Korea | Administrative claims |
| DCMC | South Korea | Electronic health records |

All databases standardized to OMOP CDM v5.3



# Live demo of CHARYBDIS

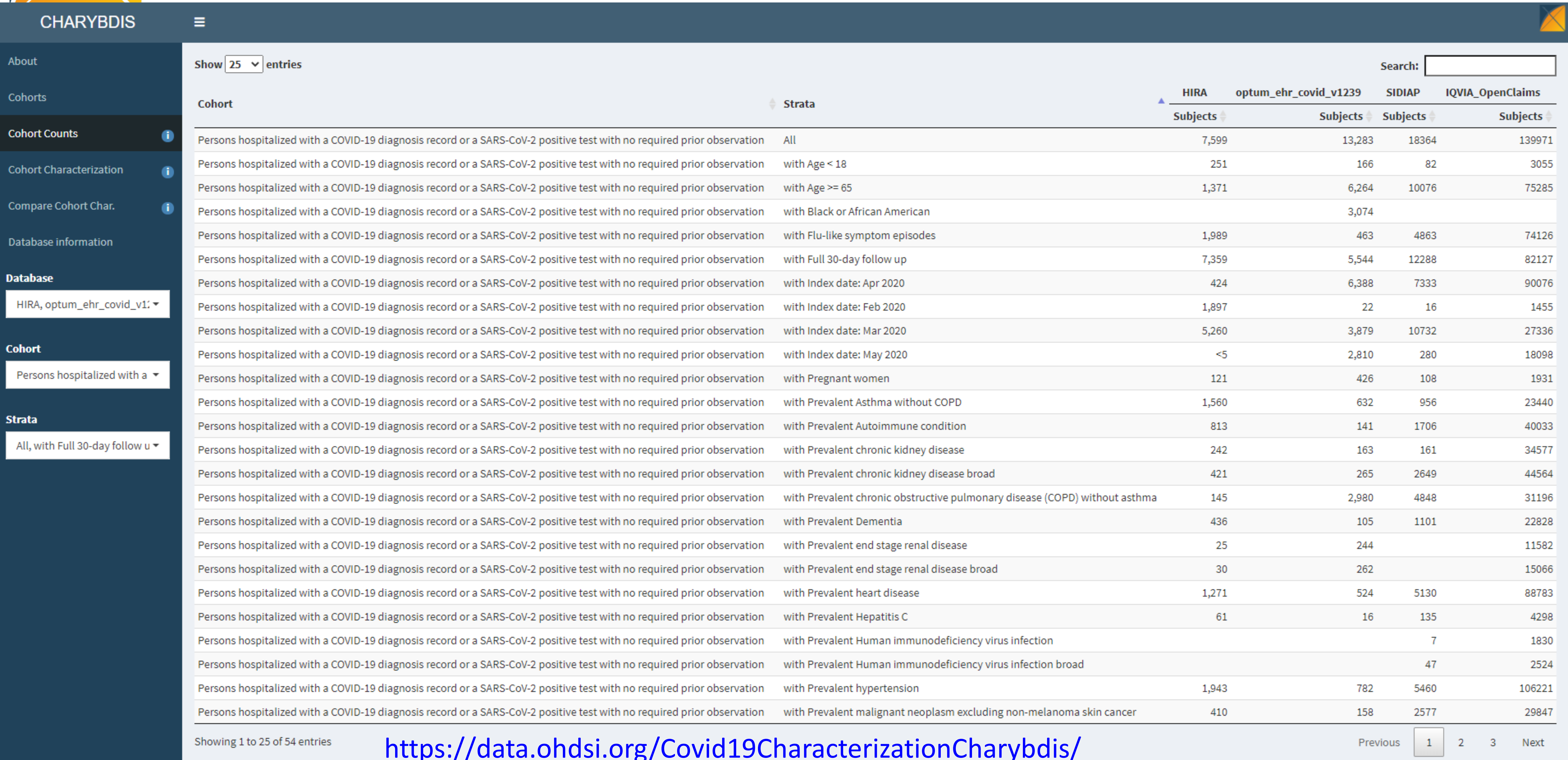


| Cohort | Strata | CDM_Premier_COVID_v1240 | HIRA | optum_ehr_covid_v1239 | SIDIAP | STARR-OMOP | TRDW | IQVIA_OpenClaims | CUIMC |
|---|---|---|---|---|---|---|---|---|---|
| | | Subjects | Subjects | Subjects | Subjects | Subjects | Subjects | Subjects | Subjects |
| Persons tested for SARS-CoV-2 with no required prior observation | All | 219,230 | 230,268 | 411,580 | 150,187 | 56,881 | 6,950 | 783214 | 22094 |
| Persons tested for SARS-CoV-2 with at least 365d prior observation | All | 1,289 | 230,268 | 355,014 | 148,468 | 39,877 | 3,719 | 739518 | 18053 |
| Persons with a COVID-19 diagnosis or a SARS-CoV-2 positive test with no required prior observation | All | 66,132 | 7,603 | 45,508 | 124,221 | 4,788 | 1,250 | 493949 | 10437 |
| Persons tested with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | All | 21,503 | 6,013 | 43,386 | 42,325 | 4,095 | 1,097 | 74793 | 7998 |
| Persons tested positive for SARS-CoV-2 with no required prior observation | All | | | 42,909 | 37,975 | 1,880 | 1,035 | | 6959 |
| Persons with a COVID-19 diagnosis or a SARS-CoV-2 positive test with at least 365d prior observation | All | 194 | 7,603 | 37,880 | 122,058 | 3,328 | 664 | 466191 | 8519 |
| Persons tested with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with at least 365d prior observation | All | 63 | 6,013 | 36,048 | 41,916 | 2,741 | 574 | 70301 | 6497 |
| Persons tested positive for SARS-CoV-2 with at least 365d prior observation | All | | | 35,624 | 37,604 | 902 | 520 | | 5625 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | All | 36,019 | 7,599 | 13,283 | 18,364 | 744 | 326 | 139971 | 3439 |
| Persons hospitalized with a SARS-CoV-2 positive test with no required prior observation | All | | | 12,451 | 13,644 | 128 | 232 | | 3075 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with at least 365d prior observation | All | 132 | 7,599 | 10,534 | 18,197 | 615 | 186 | 133091 | 2600 |
| Persons hospitalized with a SARS-CoV-2 positive test with at least 365d prior observation | All | | | 9,841 | 13,520 | 86 | 140 | | 2344 |
| Persons hospitalized and requiring intensive services with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | All | 8,373 | 130 | 1,719 | | 62 | 102 | 15184 | 86 |
| Persons hospitalized and requiring intensive services with a SARS-CoV-2 positive test with no required prior observation | All | | | 1,611 | | 19 | 73 | | 58 |
| Persons hospitalized and requiring intensive services with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with at least 365d prior observation | All | 28 | 130 | 1,345 | | 46 | 40 | 14633 | 56 |
| Persons hospitalized and requiring intensive services with a SARS-CoV-2 positive test with at least 365d prior observation | All | | | 1,253 | | 12 | 31 | | 40 |

Showing 1 to 16 of 16 entries

https://data.ohdsi.org/Covid19CharacterizationCharybdis/

# Live demo of CHARYBDIS



**Database**: HIRA, optum_ehr_covid_v1:

**Cohort**: Persons hospitalized with a

**Strata**: All, with Full 30-day follow u

Show 25 entries

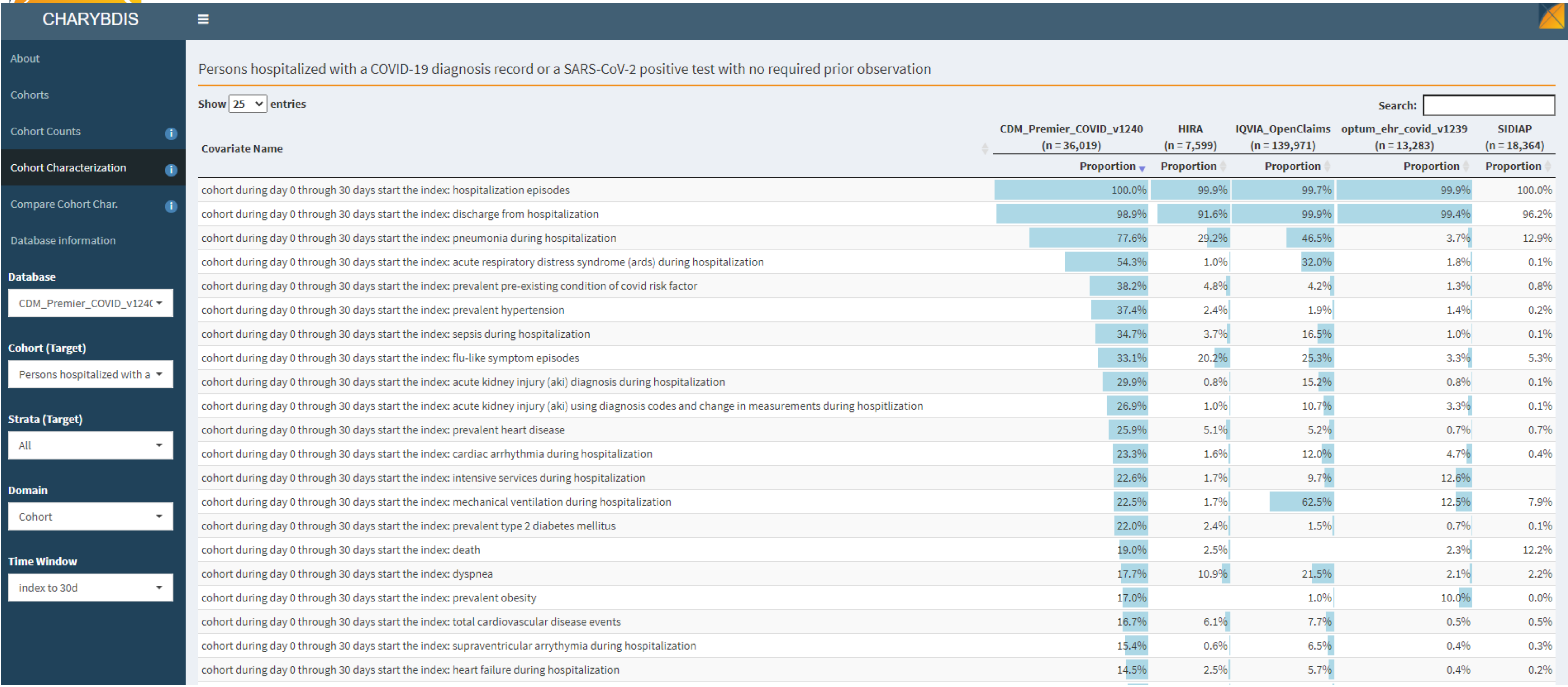
Search:

| Cohort | Strata | HIRA | optum_ehr_covid_v1239 | SIDIAP | IQVIA_OpenClaims |
|---|---|---|---|---|---|
| | | Subjects | Subjects | Subjects | Subjects |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | All | 7,599 | 13,283 | 18364 | 139971 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Age < 18 | 251 | 166 | 82 | 3055 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Age >= 65 | 1,371 | 6,264 | 10076 | 75285 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Black or African American | | 3,074 | | |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Flu-like symptom episodes | 1,989 | 463 | 4863 | 74126 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Full 30-day follow up | 7,359 | 5,544 | 12288 | 82127 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Index date: Apr 2020 | 424 | 6,388 | 7333 | 90076 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Index date: Feb 2020 | 1,897 | 22 | 16 | 1455 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Index date: Mar 2020 | 5,260 | 3,879 | 10732 | 27336 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Index date: May 2020 | <5 | 2,810 | 280 | 18098 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Pregnant women | 121 | 426 | 108 | 1931 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Prevalent Asthma without COPD | 1,560 | 632 | 956 | 23440 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Prevalent Autoimmune condition | 813 | 141 | 1706 | 40033 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Prevalent chronic kidney disease | 242 | 163 | 161 | 34577 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Prevalent chronic kidney disease broad | 421 | 265 | 2649 | 44564 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Prevalent chronic obstructive pulmonary disease (COPD) without asthma | 145 | 2,980 | 4848 | 31196 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Prevalent Dementia | 436 | 105 | 1101 | 22828 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Prevalent end stage renal disease | 25 | 244 | | 11582 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Prevalent end stage renal disease broad | 30 | 262 | | 15066 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Prevalent heart disease | 1,271 | 524 | 5130 | 88783 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Prevalent Hepatitis C | 61 | 16 | 135 | 4298 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Prevalent Human immunodeficiency virus infection | | | 7 | 1830 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Prevalent Human immunodeficiency virus infection broad | | | 47 | 2524 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Prevalent hypertension | 1,943 | 782 | 5460 | 106221 |
| Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation | with Prevalent malignant neoplasm excluding non-melanoma skin cancer | 410 | 158 | 2577 | 29847 |

Showing 1 to 25 of 54 entries

https://data.ohdsi.org/Covid19CharacterizationCharybdis/

# Live demo of CHARYBDIS



**Database**: CDM_Premier_COVID_v124(

**Cohort (Target)**: Persons hospitalized with a

**Strata (Target)**: All

**Domain**: Cohort

**Time Window**: index to 30d

Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation

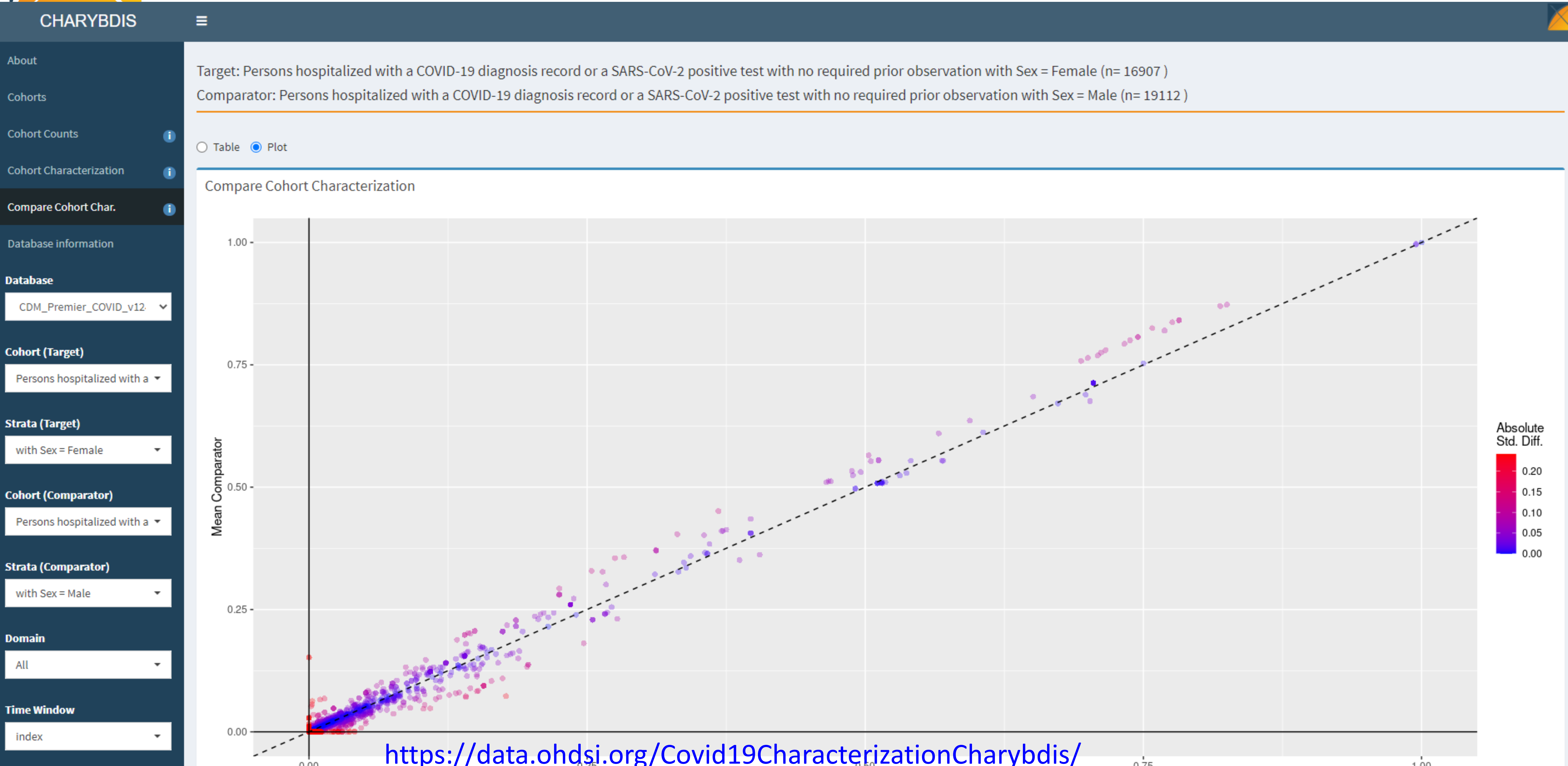
Show 25 entries

Search:

| Covariate Name | CDM_Premier_COVID_v1240 (n = 36,019) | HIRA (n = 7,599) | IQVIA_OpenClaims (n = 139,971) | optum_ehr_covid_v1239 (n = 13,283) | SIDIAP (n = 18,364) |
|---|---|---|---|---|---|
| | Proportion | Proportion | Proportion | Proportion | Proportion |
| cohort during day 0 through 30 days start the index: hospitalization episodes | 100.0% | 99.9% | 99.7% | 99.9% | 100.0% |
| cohort during day 0 through 30 days start the index: discharge from hospitalization | 98.9% | 91.6% | 99.9% | 99.4% | 96.2% |
| cohort during day 0 through 30 days start the index: pneumonia during hospitalization | 77.6% | 29.2% | 46.5% | 3.7% | 12.9% |
| cohort during day 0 through 30 days start the index: acute respiratory distress syndrome (ards) during hospitalization | 54.3% | 1.0% | 32.0% | 1.8% | 0.1% |
| cohort during day 0 through 30 days start the index: prevalent pre-existing condition of covid risk factor | 38.2% | 4.8% | 4.2% | 1.3% | 0.8% |
| cohort during day 0 through 30 days start the index: prevalent hypertension | 37.4% | 2.4% | 1.9% | 1.4% | 0.2% |
| cohort during day 0 through 30 days start the index: sepsis during hospitalization | 34.7% | 3.7% | 16.5% | 1.0% | 0.1% |
| cohort during day 0 through 30 days start the index: flu-like symptom episodes | 33.1% | 20.2% | 25.3% | 3.3% | 5.3% |
| cohort during day 0 through 30 days start the index: acute kidney injury (aki) diagnosis during hospitalization | 29.9% | 0.8% | 15.2% | 0.8% | 0.1% |
| cohort during day 0 through 30 days start the index: acute kidney injury (aki) using diagnosis codes and change in measurements during hospitlization | 26.9% | 1.0% | 10.7% | 3.3% | 0.1% |
| cohort during day 0 through 30 days start the index: prevalent heart disease | 25.9% | 5.1% | 5.2% | 0.7% | 0.7% |
| cohort during day 0 through 30 days start the index: cardiac arrhythmia during hospitalization | 23.3% | 1.6% | 12.0% | 4.7% | 0.4% |
| cohort during day 0 through 30 days start the index: intensive services during hospitalization | 22.6% | 1.7% | 9.7% | 12.6% | |
| cohort during day 0 through 30 days start the index: mechanical ventilation during hospitalization | 22.5% | 1.7% | 62.5% | 12.5% | 7.9% |
| cohort during day 0 through 30 days start the index: prevalent type 2 diabetes mellitus | 22.0% | 2.4% | 1.5% | 0.7% | 0.1% |
| cohort during day 0 through 30 days start the index: death | 19.0% | 2.5% | | 2.3% | 12.2% |
| cohort during day 0 through 30 days start the index: dyspnea | 17.7% | 10.9% | 21.5% | 2.1% | 2.2% |
| cohort during day 0 through 30 days start the index: prevalent obesity | 17.0% | | 1.0% | 10.0% | 0.0% |
| cohort during day 0 through 30 days start the index: total cardiovascular disease events | 16.7% | 6.1% | 7.7% | 0.5% | 0.5% |
| cohort during day 0 through 30 days start the index: supraventricular arrythymia during hospitalization | 15.4% | 0.6% | 6.5% | 0.4% | 0.3% |
| cohort during day 0 through 30 days start the index: heart failure during hospitalization | 14.5% | 2.5% | 5.7% | 0.4% | 0.2% |

https://data.ohdsi.org/Covid19CharacterizationCharybdis/

# Live demo of CHARYBDIS



Database: CDM_Premier_COVID_v12

Cohort (Target): Persons hospitalized with a

Strata (Target): with Sex = Female

Cohort (Comparator): Persons hospitalized with a

Strata (Comparator): with Sex = Male

Domain: All

Time Window: index

Target: Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation with Sex = Female (n= 16907 )

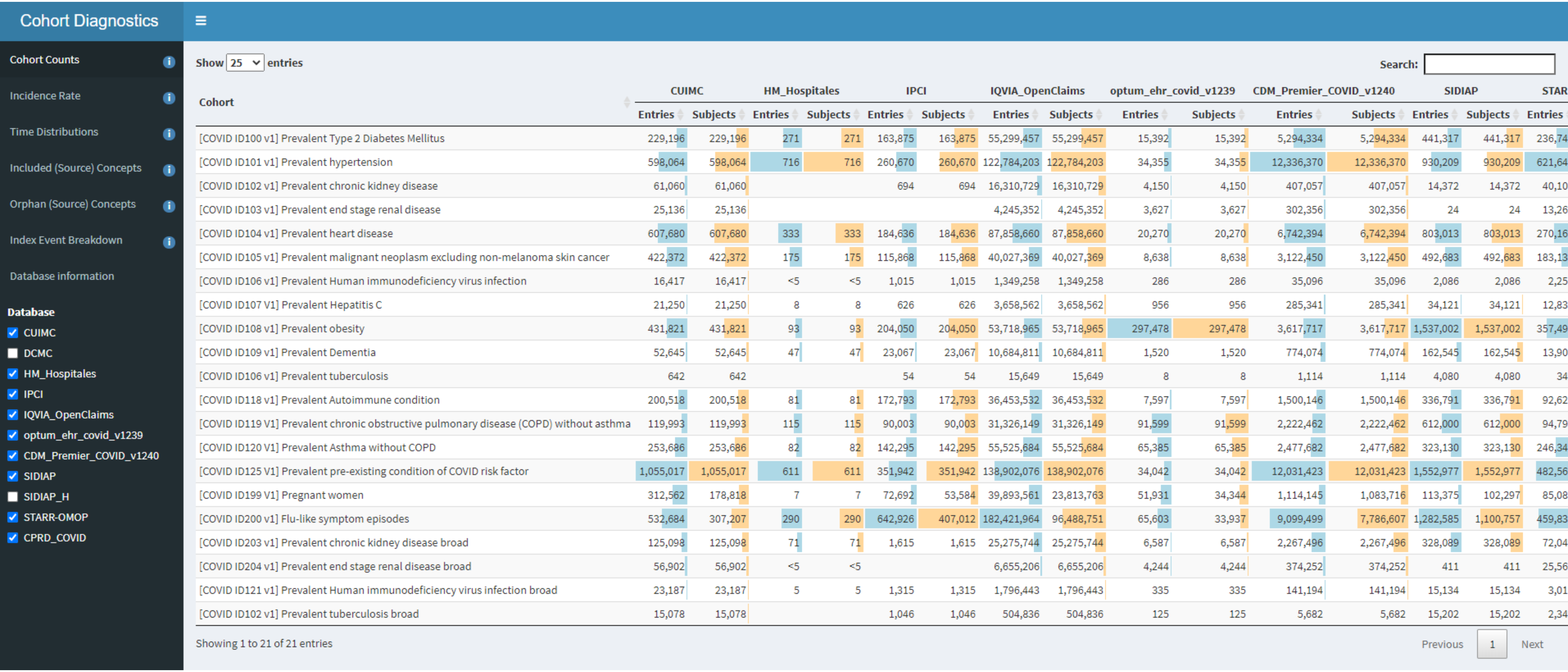
Comparator: Persons hospitalized with a COVID-19 diagnosis record or a SARS-CoV-2 positive test with no required prior observation with Sex = Male (n= 19112 )

○ Table ● Plot

Compare Cohort Characterization

https://data.ohdsi.org/Covid19CharacterizationCharybdis/

# Live demo of CHARYBDIS cohort diagnostics



**Database**

- ☑ CUIMC
- ☐ DCMC
- ☑ HM_Hospitales
- ☑ IPCI
- ☑ IQVIA_OpenClaims
- ☑ optum_ehr_covid_v1239
- ☑ CDM_Premier_COVID_v1240
- ☑ SIDIAP
- ☐ SIDIAP_H
- ☑ STARR-OMOP
- ☑ CPRD_COVID

Show 25 entries

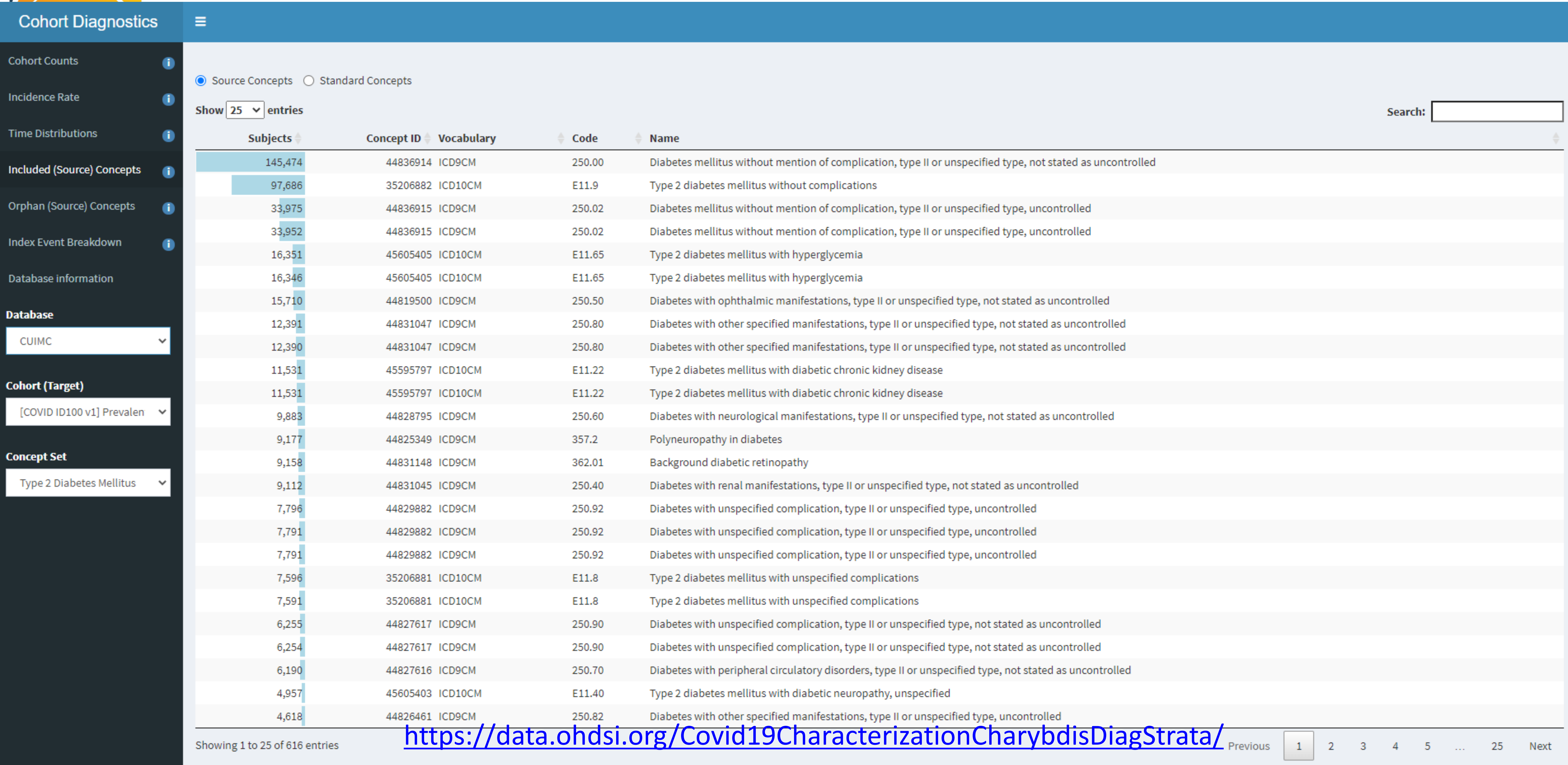
Search:

| Cohort | CUIMC | | HM_Hospitales | | IPCI | | IQVIA_OpenClaims | | optum_ehr_covid_v1239 | | CDM_Premier_COVID_v1240 | | SIDIAP | | STAR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Entries | Subjects | Entries | Subjects | Entries | Subjects | Entries | Subjects | Entries | Subjects | Entries | Subjects | Entries | Subjects | Entries |
| [COVID ID100 v1] Prevalent Type 2 Diabetes Mellitus | 229,196 | 229,196 | 271 | 271 | 163,875 | 163,875 | 55,299,457 | 55,299,457 | 15,392 | 15,392 | 5,294,334 | 5,294,334 | 441,317 | 441,317 | 236,74 |
| [COVID ID101 v1] Prevalent hypertension | 598,064 | 598,064 | 716 | 716 | 260,670 | 260,670 | 122,784,203 | 122,784,203 | 34,355 | 34,355 | 12,336,370 | 12,336,370 | 930,209 | 930,209 | 621,64 |
| [COVID ID102 v1] Prevalent chronic kidney disease | 61,060 | 61,060 | | | 694 | 694 | 16,310,729 | 16,310,729 | 4,150 | 4,150 | 407,057 | 407,057 | 14,372 | 14,372 | 40,10 |
| [COVID ID103 v1] Prevalent end stage renal disease | 25,136 | 25,136 | | | | | 4,245,352 | 4,245,352 | 3,627 | 3,627 | 302,356 | 302,356 | 24 | 24 | 13,26 |
| [COVID ID104 v1] Prevalent heart disease | 607,680 | 607,680 | 333 | 333 | 184,636 | 184,636 | 87,858,660 | 87,858,660 | 20,270 | 20,270 | 6,742,394 | 6,742,394 | 803,013 | 803,013 | 270,16 |
| [COVID ID105 v1] Prevalent malignant neoplasm excluding non-melanoma skin cancer | 422,372 | 422,372 | 175 | 175 | 115,868 | 115,868 | 40,027,369 | 40,027,369 | 8,638 | 8,638 | 3,122,450 | 3,122,450 | 492,683 | 492,683 | 183,13 |
| [COVID ID106 v1] Prevalent Human immunodeficiency virus infection | 16,417 | 16,417 | <5 | <5 | 1,015 | 1,015 | 1,349,258 | 1,349,258 | 286 | 286 | 35,096 | 35,096 | 2,086 | 2,086 | 2,25 |
| [COVID ID107 V1] Prevalent Hepatitis C | 21,250 | 21,250 | 8 | 8 | 626 | 626 | 3,658,562 | 3,658,562 | 956 | 956 | 285,341 | 285,341 | 34,121 | 34,121 | 12,83 |
| [COVID ID108 v1] Prevalent obesity | 431,821 | 431,821 | 93 | 93 | 204,050 | 204,050 | 53,718,965 | 53,718,965 | 297,478 | 297,478 | 3,617,717 | 3,617,717 | 1,537,002 | 1,537,002 | 357,49 |
| [COVID ID109 v1] Prevalent Dementia | 52,645 | 52,645 | 47 | 47 | 23,067 | 23,067 | 10,684,811 | 10,684,811 | 1,520 | 1,520 | 774,074 | 774,074 | 162,545 | 162,545 | 13,90 |
| [COVID ID106 v1] Prevalent tuberculosis | 642 | 642 | | | 54 | 54 | 15,649 | 15,649 | 8 | 8 | 1,114 | 1,114 | 4,080 | 4,080 | 34 |
| [COVID ID118 v1] Prevalent Autoimmune condition | 200,518 | 200,518 | 81 | 81 | 172,793 | 172,793 | 36,453,532 | 36,453,532 | 7,597 | 7,597 | 1,500,146 | 1,500,146 | 336,791 | 336,791 | 92,62 |
| [COVID ID119 V1] Prevalent chronic obstructive pulmonary disease (COPD) without asthma | 119,993 | 119,993 | 115 | 115 | 90,003 | 90,003 | 31,326,149 | 31,326,149 | 91,599 | 91,599 | 2,222,462 | 2,222,462 | 612,000 | 612,000 | 94,79 |
| [COVID ID120 V1] Prevalent Asthma without COPD | 253,686 | 253,686 | 82 | 82 | 142,295 | 142,295 | 55,525,684 | 55,525,684 | 65,385 | 65,385 | 2,477,682 | 2,477,682 | 323,130 | 323,130 | 246,34 |
| [COVID ID125 V1] Prevalent pre-existing condition of COVID risk factor | 1,055,017 | 1,055,017 | 611 | 611 | 351,942 | 351,942 | 138,902,076 | 138,902,076 | 34,042 | 34,042 | 12,031,423 | 12,031,423 | 1,552,977 | 1,552,977 | 482,56 |
| [COVID ID199 V1] Pregnant women | 312,562 | 178,818 | 7 | 7 | 72,692 | 53,584 | 39,893,561 | 23,813,763 | 51,931 | 34,344 | 1,114,145 | 1,083,716 | 113,375 | 102,297 | 85,08 |
| [COVID ID200 v1] Flu-like symptom episodes | 532,684 | 307,207 | 290 | 290 | 642,926 | 407,012 | 182,421,964 | 96,488,751 | 65,603 | 33,937 | 9,099,499 | 7,786,607 | 1,282,585 | 1,100,757 | 459,83 |
| [COVID ID203 v1] Prevalent chronic kidney disease broad | 125,098 | 125,098 | 71 | 71 | 1,615 | 1,615 | 25,275,744 | 25,275,744 | 6,587 | 6,587 | 2,267,496 | 2,267,496 | 328,089 | 328,089 | 72,04 |
| [COVID ID204 v1] Prevalent end stage renal disease broad | 56,902 | 56,902 | <5 | <5 | | | 6,655,206 | 6,655,206 | 4,244 | 4,244 | 374,252 | 374,252 | 411 | 411 | 25,56 |
| [COVID ID121 v1] Prevalent Human immunodeficiency virus infection broad | 23,187 | 23,187 | 5 | 5 | 1,315 | 1,315 | 1,796,443 | 1,796,443 | 335 | 335 | 141,194 | 141,194 | 15,134 | 15,134 | 3,01 |
| [COVID ID102 v1] Prevalent tuberculosis broad | 15,078 | 15,078 | | | 1,046 | 1,046 | 504,836 | 504,836 | 125 | 125 | 5,682 | 5,682 | 15,202 | 15,202 | 2,34 |

Showing 1 to 21 of 21 entries

https://data.ohdsi.org/Covid19CharacterizationCharybdisDiagStrata/

# Live demo of CHARYBDIS cohort diagnostics



**Database**: CUIMC

**Cohort (Target)**: [COVID ID100 v1] Prevalen

**Concept Set**: Type 2 Diabetes Mellitus

● Source Concepts ○ Standard Concepts

Show 25 entries

Search:

| Subjects | Concept ID | Vocabulary | Code | Name |
|---|---|---|---|---|
| 145,474 | 44836914 | ICD9CM | 250.00 | Diabetes mellitus without mention of complication, type II or unspecified type, not stated as uncontrolled |
| 97,686 | 35206882 | ICD10CM | E11.9 | Type 2 diabetes mellitus without complications |
| 33,975 | 44836915 | ICD9CM | 250.02 | Diabetes mellitus without mention of complication, type II or unspecified type, uncontrolled |
| 33,952 | 44836915 | ICD9CM | 250.02 | Diabetes mellitus without mention of complication, type II or unspecified type, uncontrolled |
| 16,351 | 45605405 | ICD10CM | E11.65 | Type 2 diabetes mellitus with hyperglycemia |
| 16,346 | 45605405 | ICD10CM | E11.65 | Type 2 diabetes mellitus with hyperglycemia |
| 15,710 | 44819500 | ICD9CM | 250.50 | Diabetes with ophthalmic manifestations, type II or unspecified type, not stated as uncontrolled |
| 12,391 | 44831047 | ICD9CM | 250.80 | Diabetes with other specified manifestations, type II or unspecified type, not stated as uncontrolled |
| 12,390 | 44831047 | ICD9CM | 250.80 | Diabetes with other specified manifestations, type II or unspecified type, not stated as uncontrolled |
| 11,531 | 45595797 | ICD10CM | E11.22 | Type 2 diabetes mellitus with diabetic chronic kidney disease |
| 11,531 | 45595797 | ICD10CM | E11.22 | Type 2 diabetes mellitus with diabetic chronic kidney disease |
| 9,883 | 44828795 | ICD9CM | 250.60 | Diabetes with neurological manifestations, type II or unspecified type, not stated as uncontrolled |
| 9,177 | 44825349 | ICD9CM | 357.2 | Polyneuropathy in diabetes |
| 9,158 | 44831148 | ICD9CM | 362.01 | Background diabetic retinopathy |
| 9,112 | 44831045 | ICD9CM | 250.40 | Diabetes with renal manifestations, type II or unspecified type, not stated as uncontrolled |
| 7,796 | 44829882 | ICD9CM | 250.92 | Diabetes with unspecified complication, type II or unspecified type, uncontrolled |
| 7,791 | 44829882 | ICD9CM | 250.92 | Diabetes with unspecified complication, type II or unspecified type, uncontrolled |
| 7,791 | 44829882 | ICD9CM | 250.92 | Diabetes with unspecified complication, type II or unspecified type, uncontrolled |
| 7,596 | 35206881 | ICD10CM | E11.8 | Type 2 diabetes mellitus with unspecified complications |
| 7,591 | 35206881 | ICD10CM | E11.8 | Type 2 diabetes mellitus with unspecified complications |
| 6,255 | 44827617 | ICD9CM | 250.90 | Diabetes with unspecified complication, type II or unspecified type, not stated as uncontrolled |
| 6,254 | 44827617 | ICD9CM | 250.90 | Diabetes with unspecified complication, type II or unspecified type, not stated as uncontrolled |
| 6,190 | 44827616 | ICD9CM | 250.70 | Diabetes with peripheral circulatory disorders, type II or unspecified type, not stated as uncontrolled |
| 4,957 | 45605403 | ICD10CM | E11.40 | Type 2 diabetes mellitus with diabetic neuropathy, unspecified |
| 4,618 | 44826461 | ICD9CM | 250.82 | Diabetes with other specified manifestations, type II or unspecified type, uncontrolled |

Showing 1 to 25 of 616 entries

https://data.ohdsi.org/Covid19CharacterizationCharybdisDiagStrata/

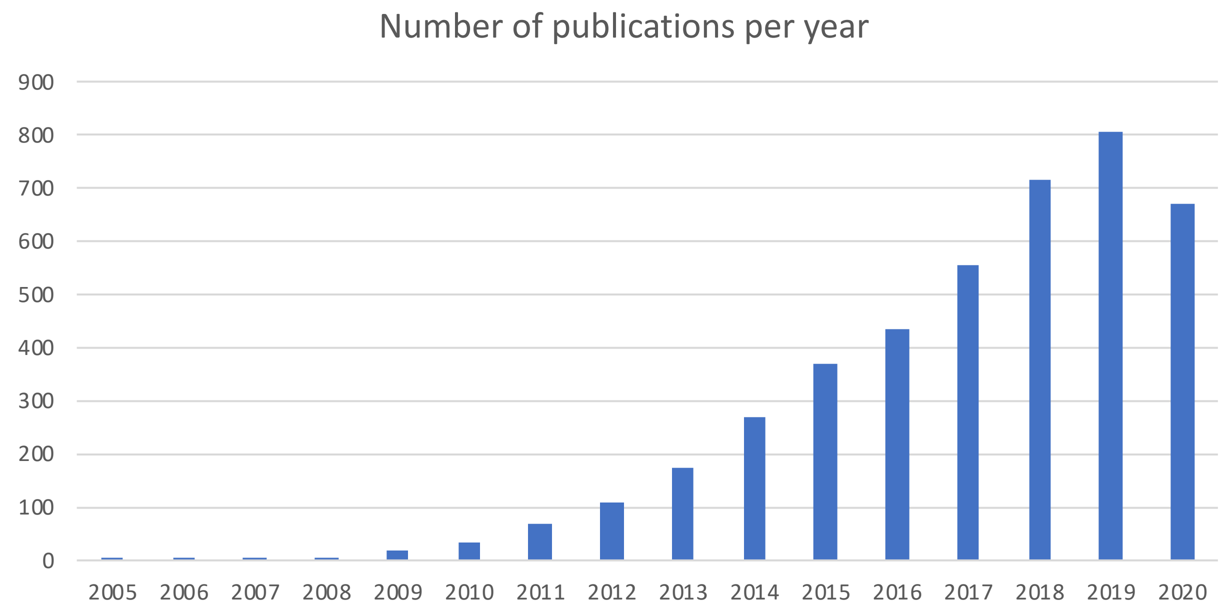
# Using Twitter to characterize the COVID disease natural history and 'life after COVID'

Juan M. Banda

www.panacealab.org

Georgia State University

# Preface: Twitter is gaining attention for health-related research since 2009



Results of PubMed Query for Twitter and Health

# Benefits of using Twitter:

1) Good population representation

2) Everybody can post and have an account

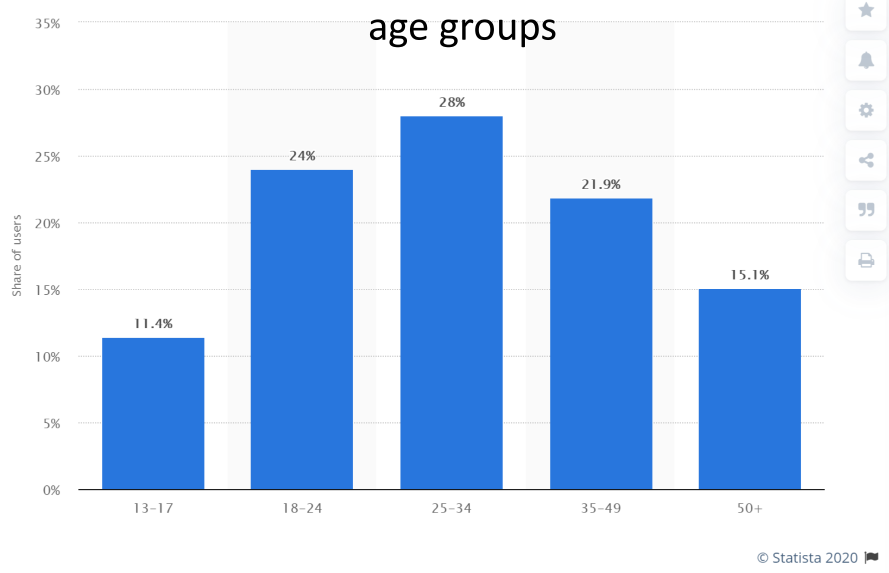
3) Anonymity = unfiltered opinions

4) Data is freely available*

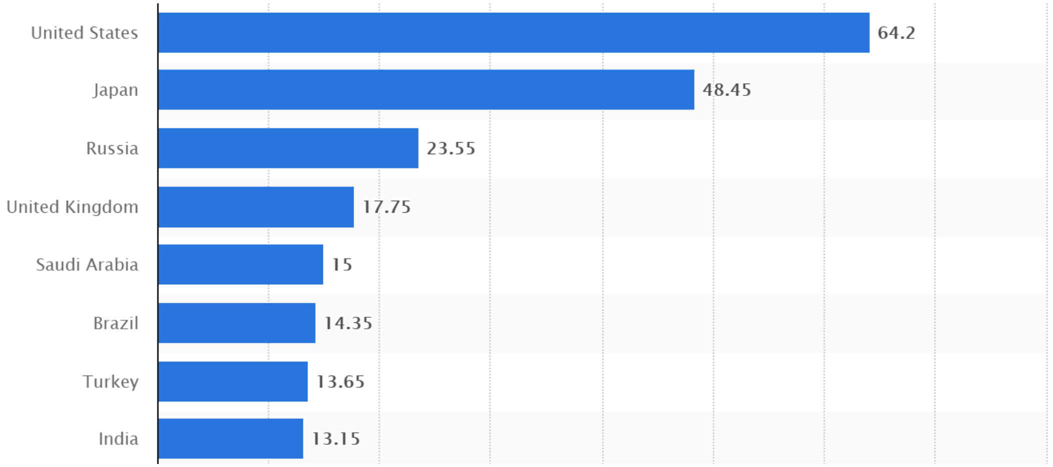
5) Tons of data generated each day (hundreds of millions of tweets get posted every day)

6) Easy filtering (hashtag usage, people mentions)

Distribution of global users over age groups



Number of users per country (millions)



https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/

# Traditional disadvantages of using Twitter:

- Messy data (plenty of misspellings, shorthand, emojis, etc.)
  - There are at least 25 different ways people misspell hydroxychloroquine

- Attribution is an issue – are people just mentioning something or did it happen to them?

- Freely available data is only a 1% sample of whole set

- Collection is hard and needs to be ongoing for days/weeks before getting considerable mass

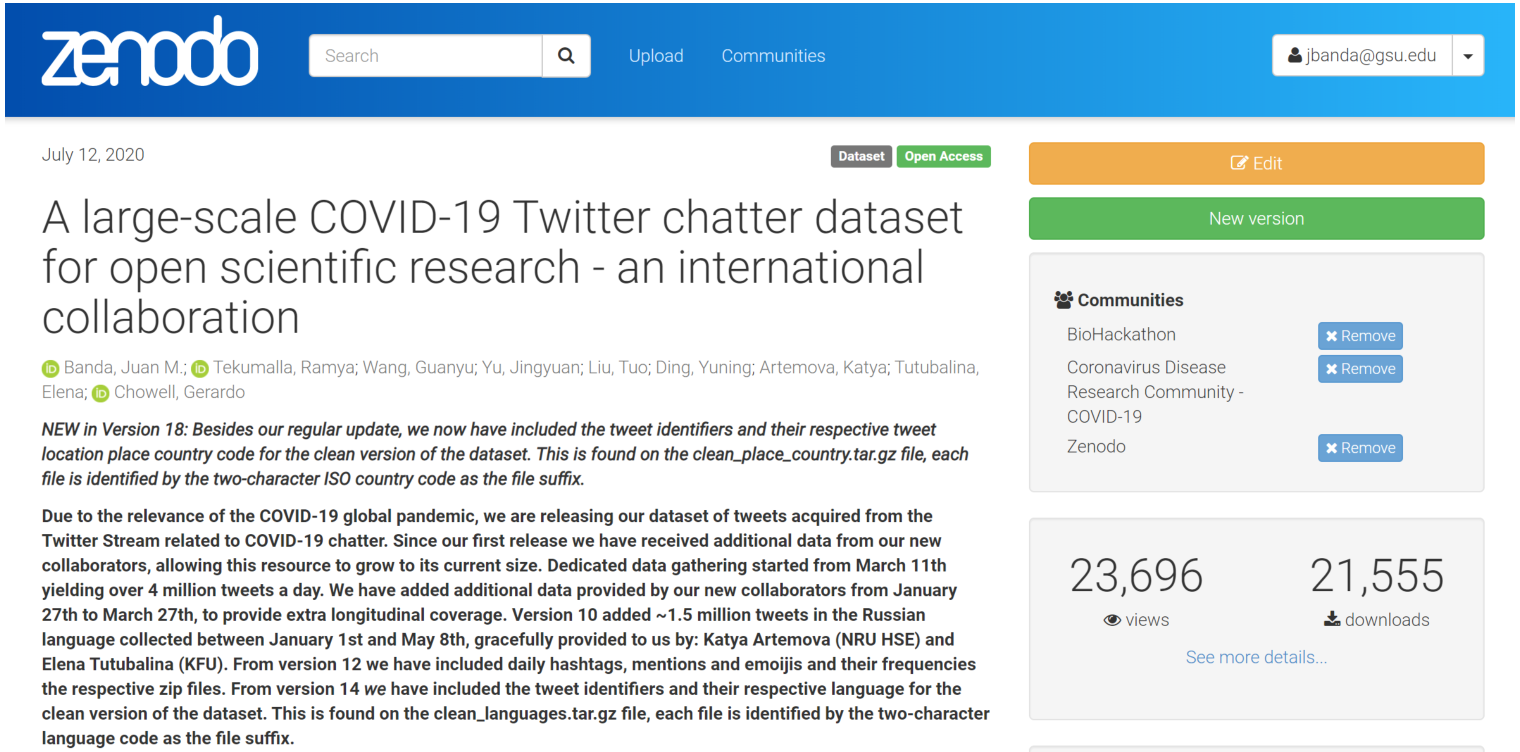
# The COVID opportunity – highly focused data

## The dataset:

- 490+ Million Tweets

- ONLY COVID related chatter is included

Longitudinal – January 27th to today… and growing



July 12, 2020

A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration

Banda, Juan M.; Tekumalla, Ramya; Wang, Guanyu; Yu, Jingyuan; Liu, Tuo; Ding, Yuning; Artemova, Katya; Tutubalina, Elena; Chowell, Gerardo

*NEW in Version 18: Besides our regular update, we now have included the tweet identifiers and their respective tweet location place country code for the clean version of the dataset. This is found on the clean_place_country.tar.gz file, each file is identified by the two-character ISO country code as the file suffix.*

**Due to the relevance of the COVID-19 global pandemic, we are releasing our dataset of tweets acquired from the Twitter Stream related to COVID-19 chatter. Since our first release we have received additional data from our new collaborators, allowing this resource to grow to its current size. Dedicated data gathering started from March 11th yielding over 4 million tweets a day. We have added additional data provided by our new collaborators from January 27th to March 27th, to provide extra longitudinal coverage. Version 10 added ~1.5 million tweets in the Russian language collected between January 1st and May 8th, gracefully provided to us by: Katya Artemova (NRU HSE) and Elena Tutubalina (KFU). From version 12 we have included daily hashtags, mentions and emoijis and their frequencies the respective zip files. From version 14 *we* have included the tweet identifiers and their respective language for the clean version of the dataset. This is found on the clean_languages.tar.gz file, each file is identified by the two-character language code as the file suffix.**

Communities

BioHackathon

Coronavirus Disease Research Community - COVID-19

Zenodo

23,696 views

21,555 downloads

Dataset: https://doi.org/10.5281/zenodo.3723939

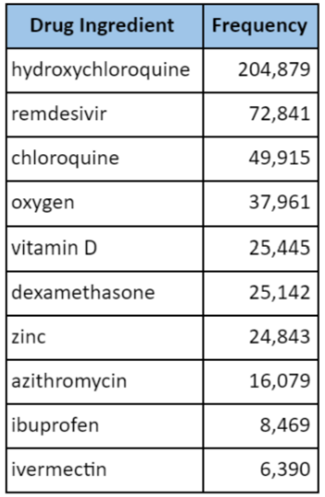
Pre-print: https://arxiv.org/abs/2004.03688

Recent additions: https://github.com/thepanacealab/covid19_twitter

# Current work: Drug characterization

- Methods to deal with misspellings and noisiness of data:

**Table 2.** **Drug ingredient mentions found**



| Drug Ingredient | Frequency |
|---|---|
| hydroxychloroquine | 204,879 |
| remdesivir | 72,841 |
| chloroquine | 49,915 |
| oxygen | 37,961 |
| vitamin D | 25,445 |
| dexamethasone | 25,142 |
| zinc | 24,843 |
| azithromycin | 16,079 |
| ibuprofen | 8,469 |
| ivermectin | 6,390 |

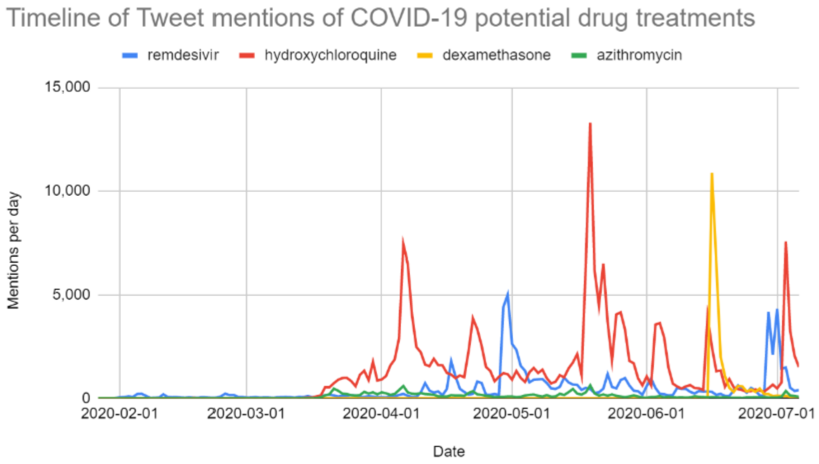


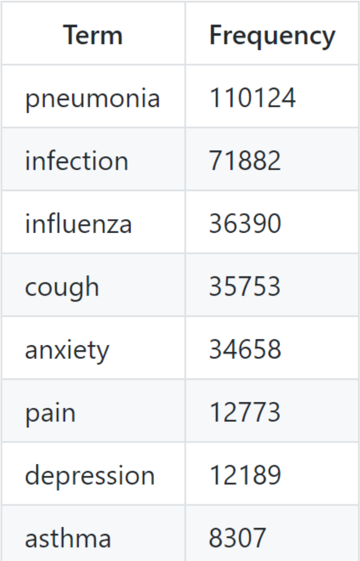
**Figure 1.** Timeline of Tweets with potential drug treatment mentions.

- Charybdis-like characterization over countries (work with Dani Prieto-Alhambra – University of Oxford)

[* https://openreview.net/forum?id=qlGPXs9FWa](https://openreview.net/forum?id=qlGPXs9FWa)

# Current Work: Symptom/condition detection

- Self-reported symptoms on Twitter vs EHR lists *
  - Can we find related symptoms both found on EHR’s (Callahan, A., Steinberg, E., Fries, J.A. et al. Estimating the efficacy of symptom-based screening for COVID-19. npj Digit. Med. 3, 95 (2020). https://doi.org/10.1038/s41746-020-0300-0) but on Twitter?



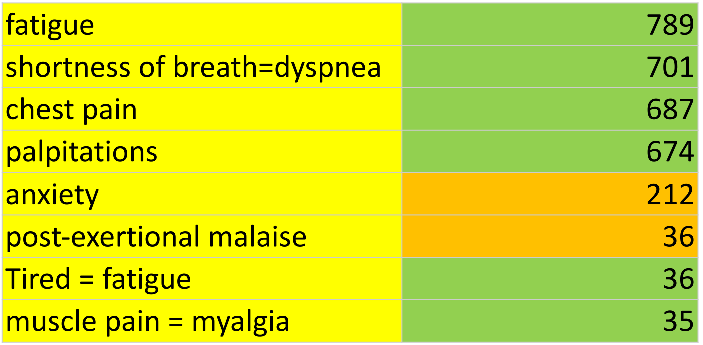
| Term | Frequency |
|---|---|
| pneumonia | 110124 |
| infection | 71882 |
| influenza | 36390 |
| cough | 35753 |
| anxiety | 34658 |
| pain | 12773 |
| depression | 12189 |
| asthma | 8307 |

* https://github.com/thepanacealab/covid19_biohackathon/tree/master/user_symptoms

# What does this lead to?

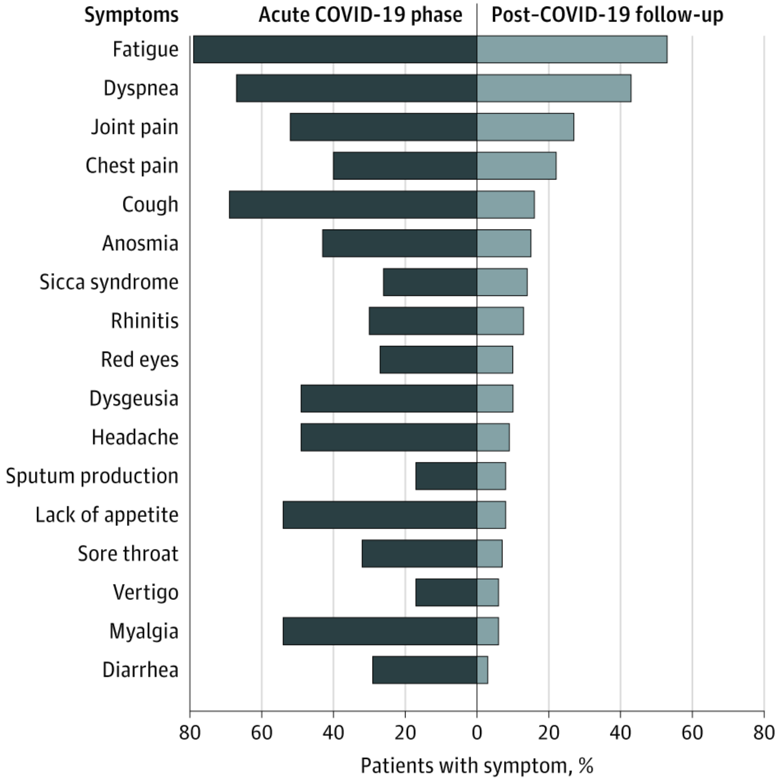
- Since we can find symptoms and drugs, we can also find people that had COVID and their symptoms after infection!
    - On-going work with Dani Prieto-Alhambra and others
        - Incorporates methods shown before + manual review by clinicians

Some very preliminary findings:



| | |
|---|---:|
| fatigue | 789 |
| shortness of breath=dyspnea | 701 |
| chest pain | 687 |
| palpitations | 674 |
| anxiety | 212 |
| post-exertional malaise | 36 |
| Tired = fatigue | 36 |
| muscle pain = myalgia | 35 |

JAMA doi:10.1001/jama.2020.12603



# The gory details:

- Technical stuff:
  - “Building tools and frameworks for large-scale social media mining: Creating data infrastructure for COVID-19 research” **dair.ai meetup 7/22:** https://www.meetup.com/dair-ai/events/271690722/

- Extended version of today’s short talk:
  - “Leveraging the OHDSI vocabulary to characterize the COVID-19 epidemic using Twitter data and NLP” **OHDSI community call 7/21**: https://www.ohdsi.org/web/wiki/doku.php?id=projects:ohdsi_community