Cohort Diagnostics

Gowtham Rao

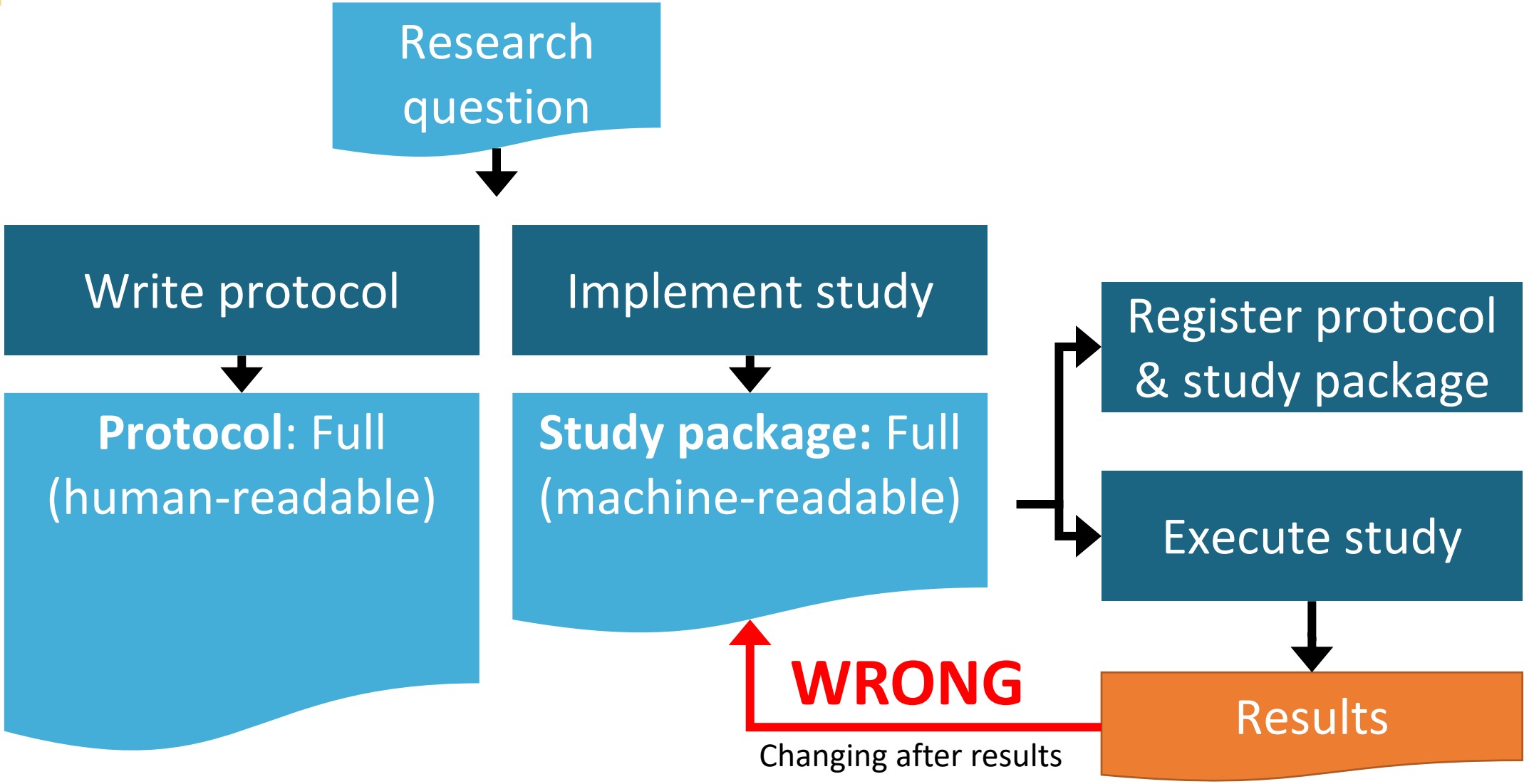
06/16/2020

Overview

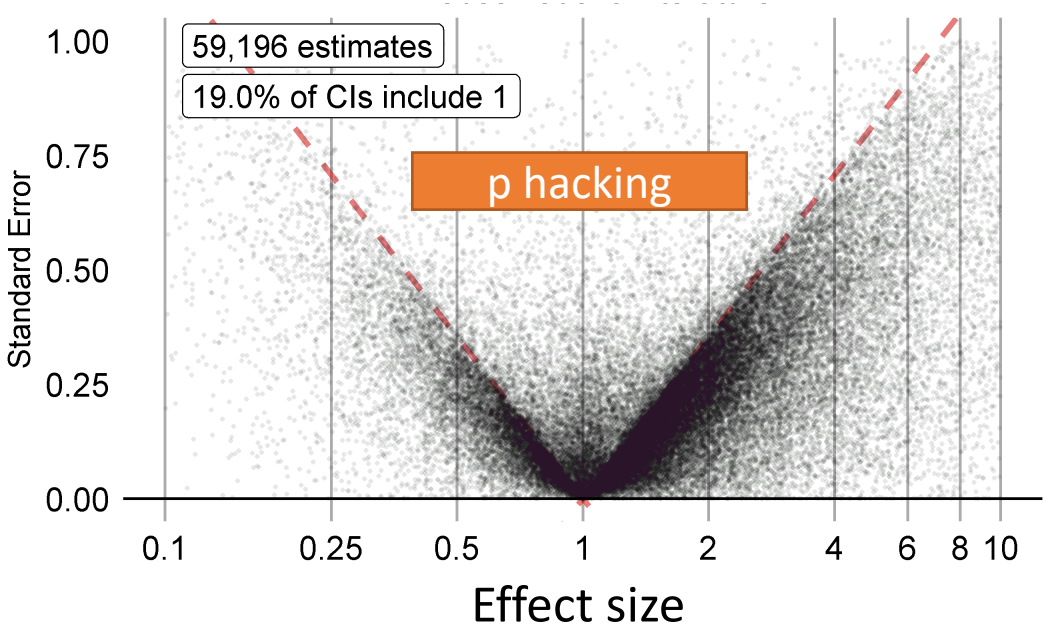
- **Motivation**

- Uses cases → software functionality

Motivation: Natural history of an observation study.



Motivation: Reduce bias/Improve confidence



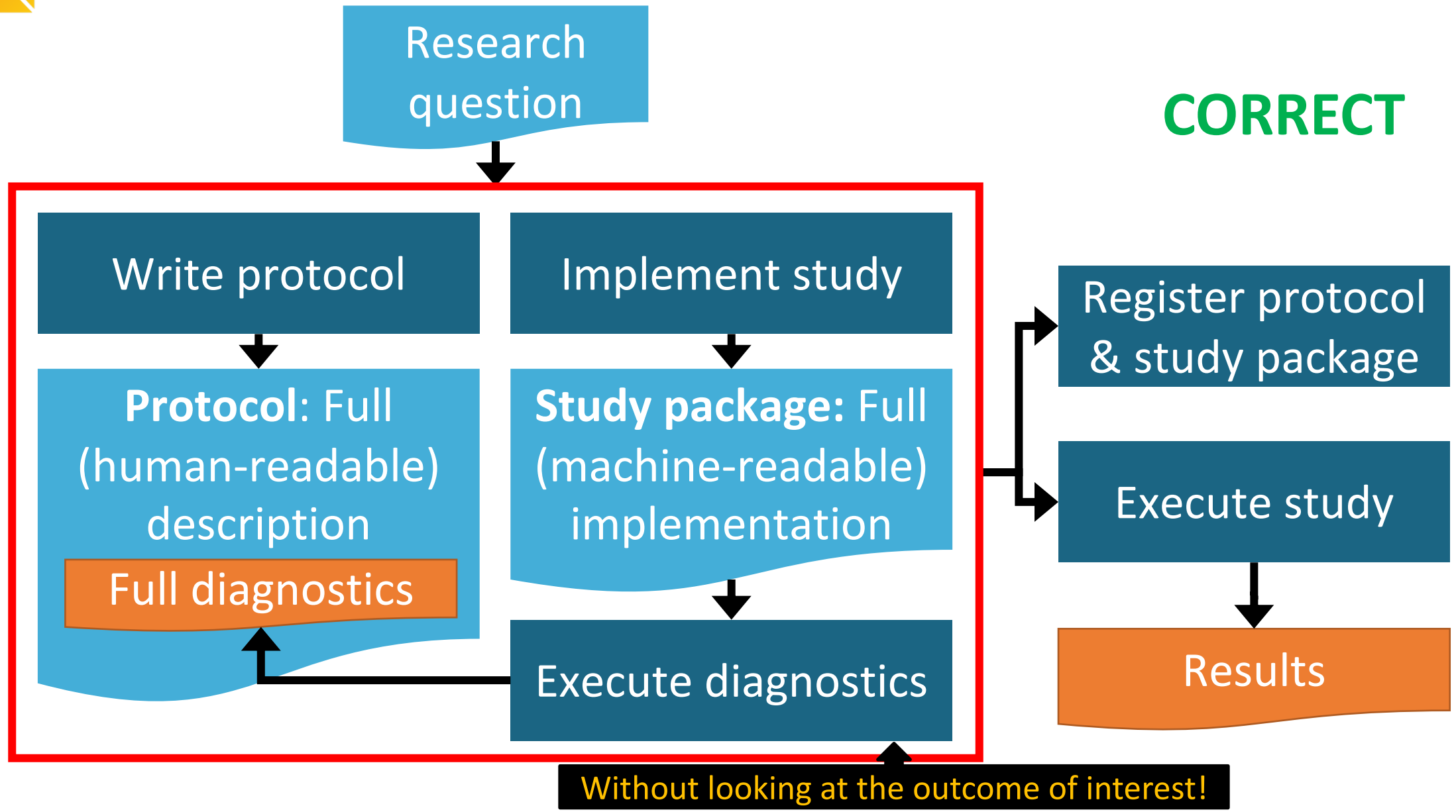
Effect size estimates in published observational studies

To make sure our study will provide a reliable answer

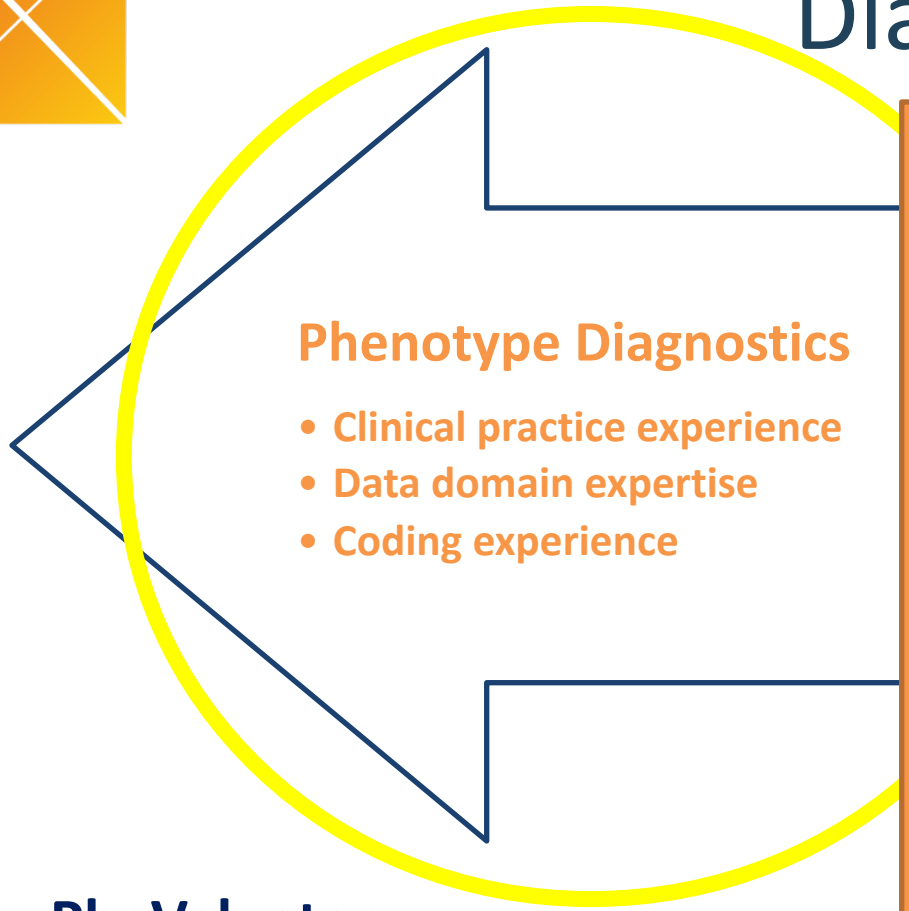
(before looking at the answer)

Changing the question/study design/dissemination based on the results leads to unreliable evidence

Diagnostics as a best practice



Diagnostics in OHDSI



Something was missing

- **Earlier in the process (phenotyping)**
  - *Before phevaluator, negative control, PS*
- **Focussed specifically on building cohorts**
  - *Feedback on impact of specification choices*
  - *Clinical description → literature review → buiding cohorts*
- **Enable conversation between investigators**
  - *Compare the impact of specification choices across data sources.*

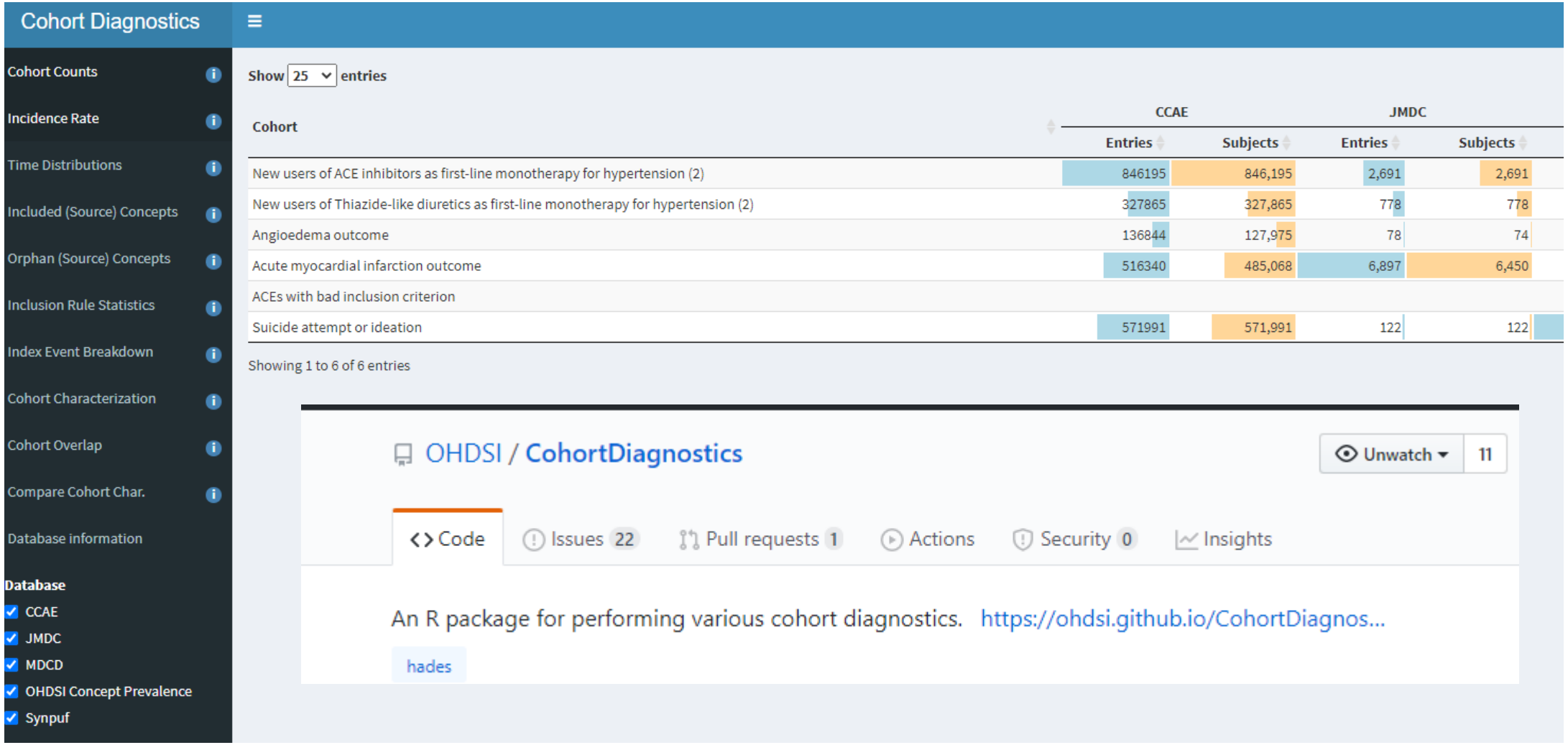
Data Quality | **Clinical Validity** | Software Validity | **Method Validity**

Overview

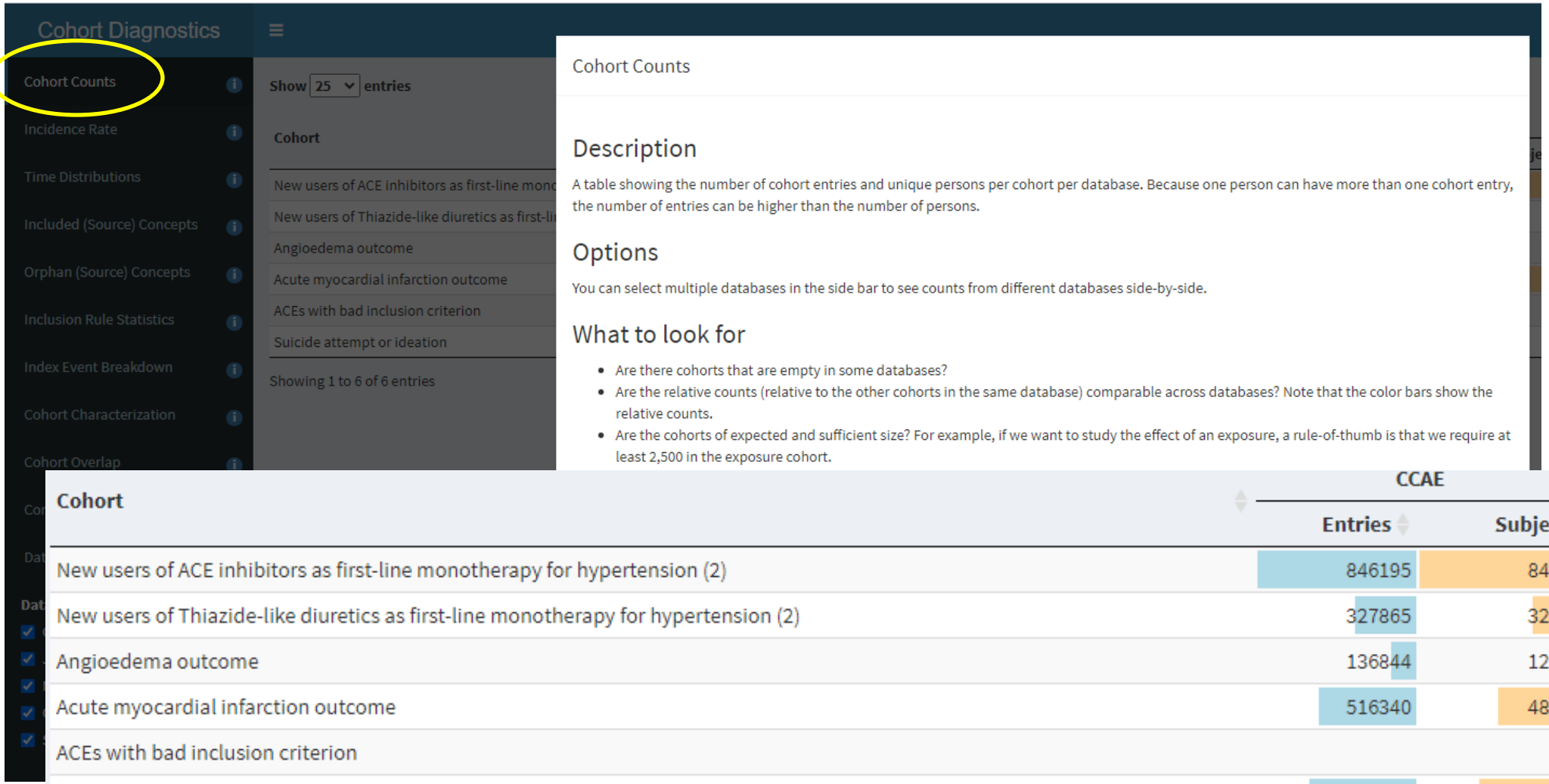
- Motivation

- **Uses cases → software functionality**

New R-package - Overview



Functionality – Cohort Counts



Cohort	CCAE	
	Entries	Subjects
New users of ACE inhibitors as first-line monotherapy for hypertension (2)	846195	846,195
New users of Thiazide-like diuretics as first-line monotherapy for hypertension (2)	327865	327,865
Angioedema outcome	136844	127,975
Acute myocardial infarction outcome	516340	485,068
ACEs with bad inclusion criterion		
Suicide attempt or ideation	571991	571,991

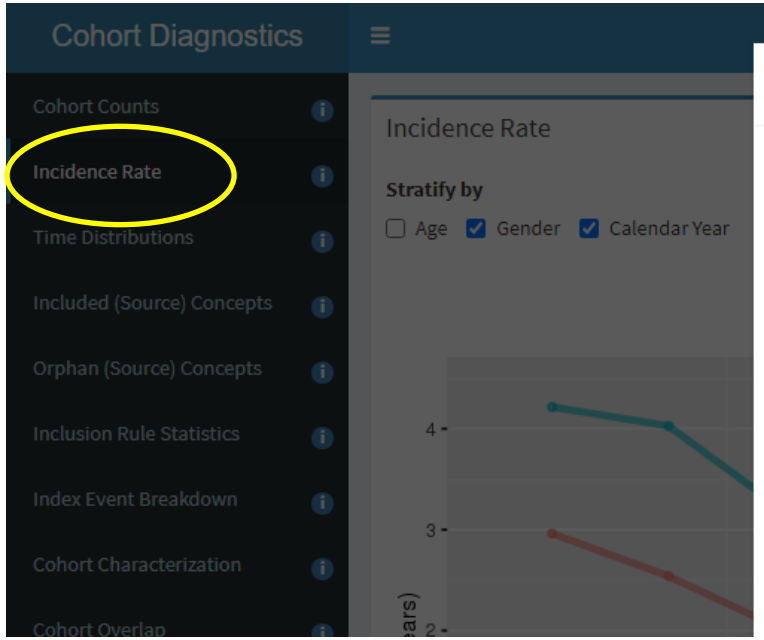
Showing 1 to 6 of 6 entries

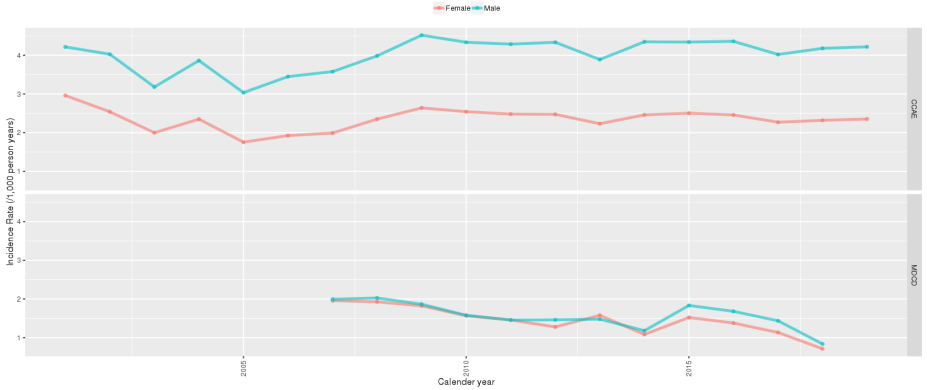
Functionality – Cohort Counts

Cohort	CPRD		IBM_CCAE		IBM_MDCD		IBM_MDCR		JMDC		OPTUM_EXTENDED_DOD		OPTUM_PANTHER	
	Entries	Subjects	Entries	Subjects	Entries	Subjects	Entries	Subjects	Entries	Subjects	Entries	Subjects	Entries	Subjects
[Cohort Diagnostics] Bacterial Pneumonia	26374	23,860	1,311,797	854,098	749,091	435,833	669,162	352,326	208,258	89,700	1,935,043	910,276	1,487,799	668511
[Cohort Diagnostics] Hospitalizations with Pneumonia			746,581	625,029	747,000	514,456	956,500	706,649	54,885	46,550	1,632,905	1,211,934	910,717	730028
[Cohort Diagnostics] Pneumonia	220467	193,475	4,136,169	3,278,670	2,157,561	1,395,845	2,068,702	1,206,277	441,396	317,061	4,369,398	3,003,301	4,654,758	3158412
[Cohort Diagnostics] Viral Pneumonia	30272	30,272	170,179	170,179	62,923	62,923	91,668	91,668	717	717	215,328	215,328	112,242	112242
[Cohort Diagnostics] Pneumonia with viral code events	30482	29,689	157,532	152,534	70,760	68,501	42,206	40,470	6,158	5,956	140,995	136,030	96,690	94099
[Cohort Diagnostics] Pneumonia without non-viral source events	202079	186,365	3,559,235	3,144,556	1,710,875	1,338,548	1,570,292	1,161,432	330,695	277,160	3,490,792	2,871,062	3,798,057	3073883

Concept Name
Pneumonia
Ornithosis with pneumonia
Idiopathic pneumonia syndrome
Fungal pneumonia
Atypical pneumonia
Acute ulcerative gastroenteritis complicating pneumonia
Acute pneumonia due to coccidioidomycosis

Functionality – Incidence Rate





Incidence Rate

Description

A graph showing the incidence rate, optionally stratified by age (in 10-year bins), gender, and calendar year.

The incidence rate is computed as 1000 * the number of people *first* entering the cohort / the number of years people were eligible to enter the cohort for the first time. The eligble person time is defined as the time when

- A person was observed in the database (based on the observation_period table).
- Had the required amount of prior observation time as specified in the cohort entry event criteria. For example, if the cohort definition requires 365 days of observation prior to cohort entry, patients are not eligble to enter the cohort in the first 365 days of their observation period, and this time is not counted in the eligible time.
- If the person enters the cohort, then only the time up to cohort entry. Because we only consider the first cohort entry, persons are no longer eligible to enter to cohort after their first entry.

Options

You can select multiple databases in the side bar to see graphs from different databases in the same plot.

Select the cohort to explore in the side bar.

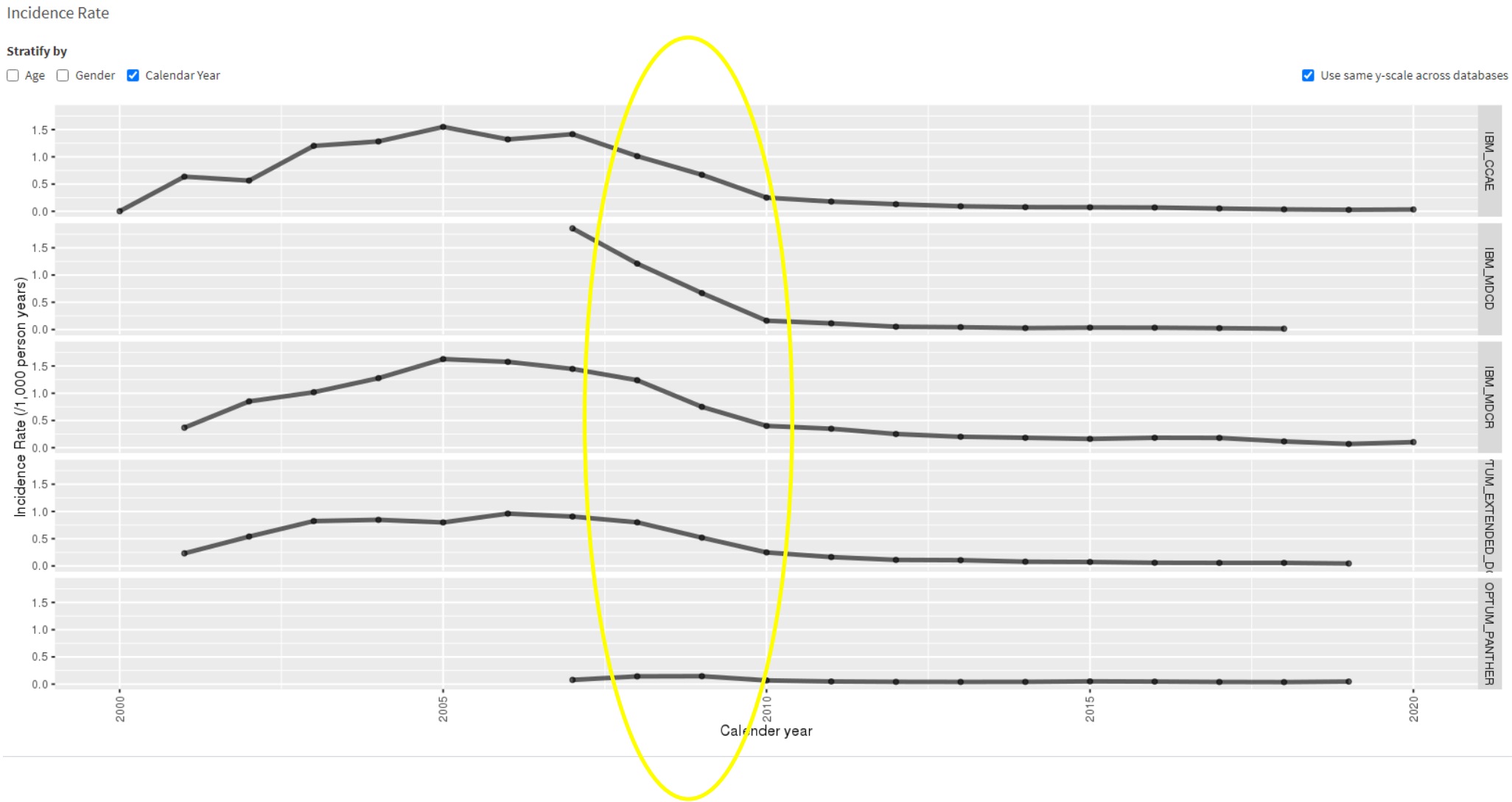
At the top of the plot, you can choose whether to stratify the data by age, gender, or calendar year.

If you move the mouse over the plot, you can see the precise value.

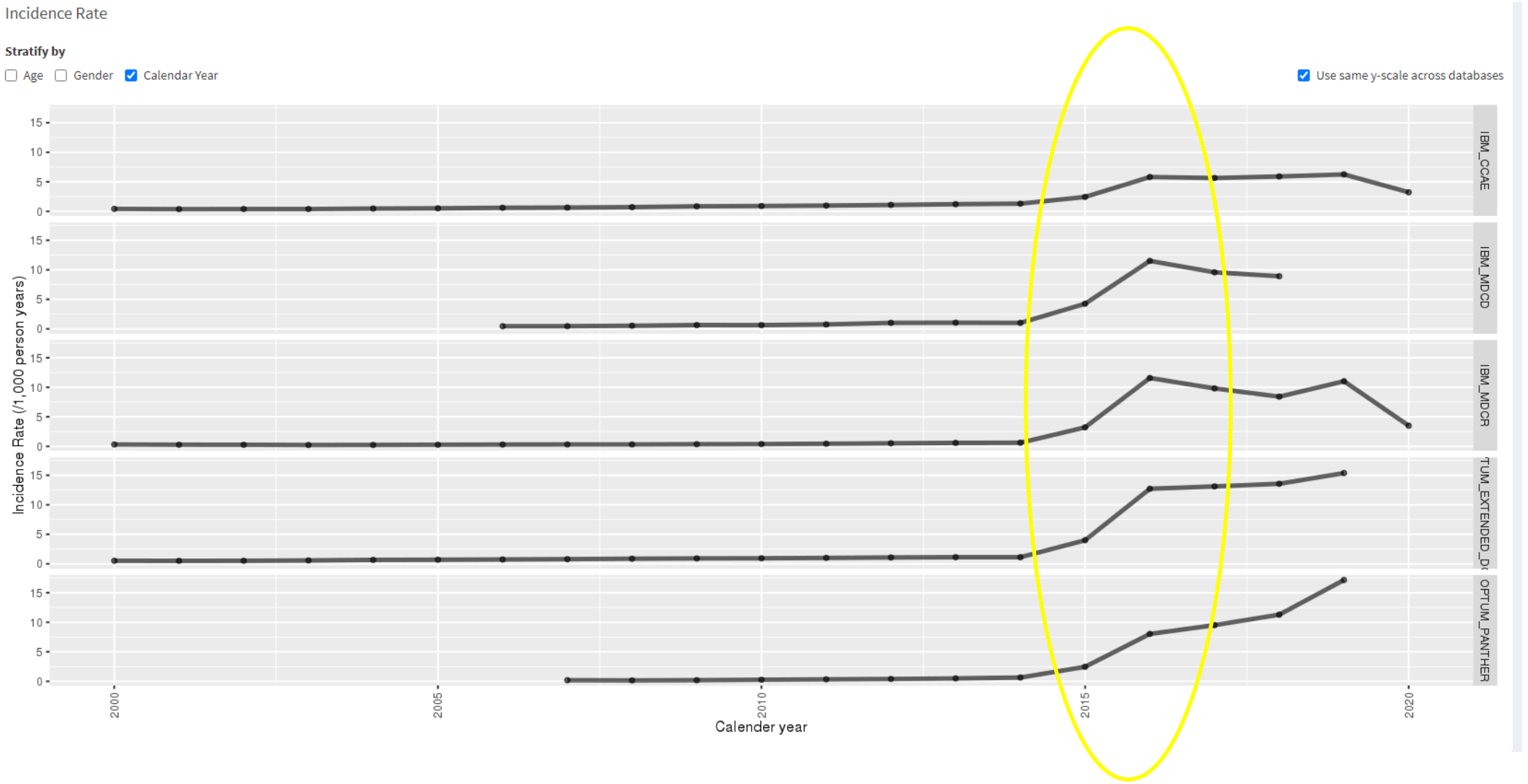
What to look for

- Are the observed incidence rates in line with expectations? For example, if we have an estimate of the population incidence based on an external source, is the incidence rate comparable to that estimate?
- Are the age and gender distributions in line with expectations? For example, are contraceptives only prescribed in women?
- Is the incidence rate stable over time? If there are sudden peaks or drops, this may indicate coding issues.

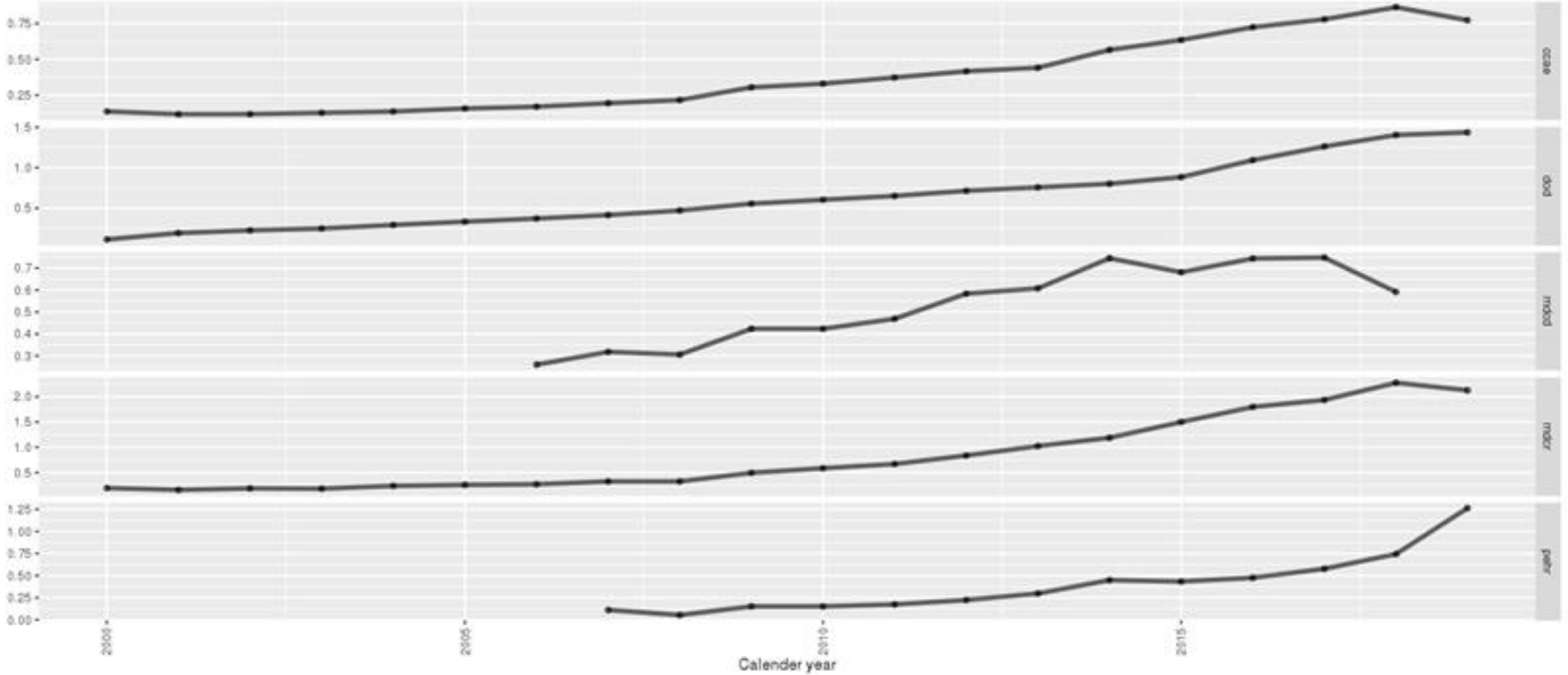
Functionality – Incidence Rate
Dramatic drop around 2008



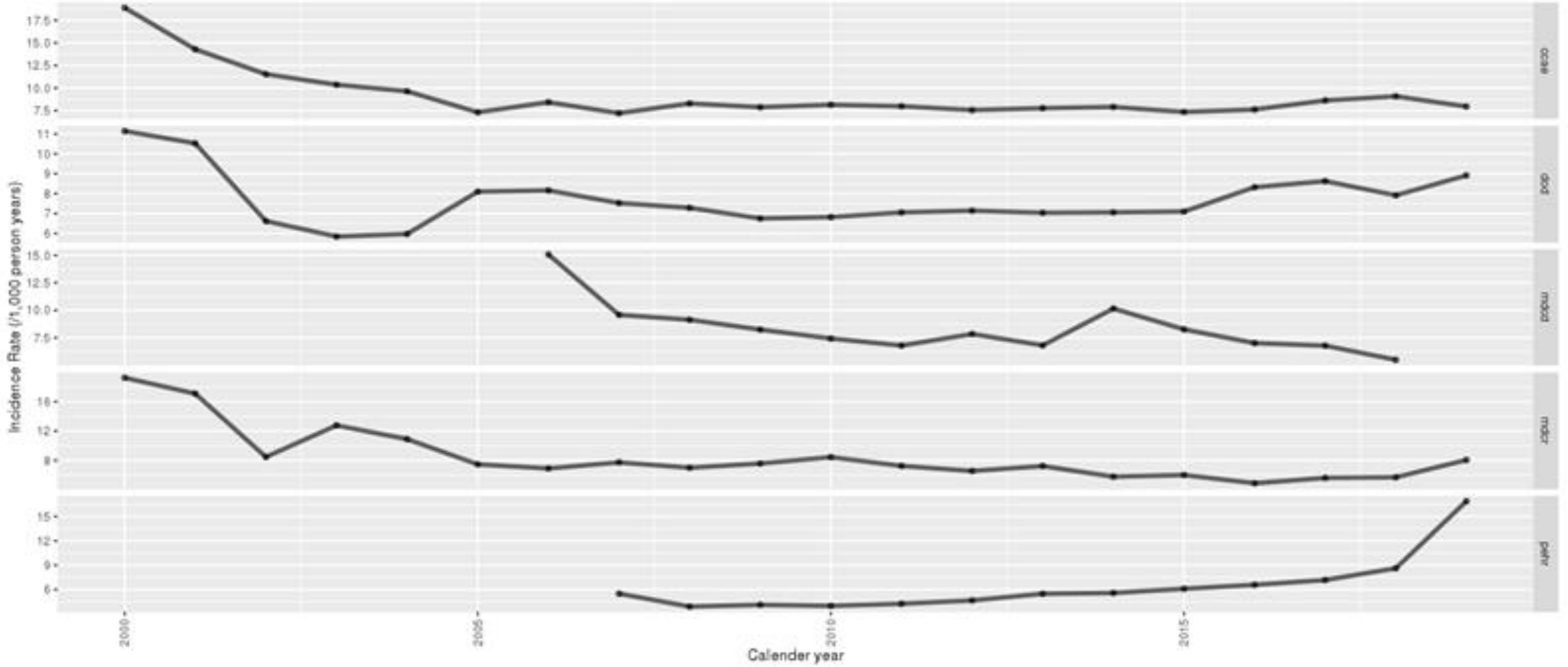
Functionality – Incidence Rate Increase around 2015



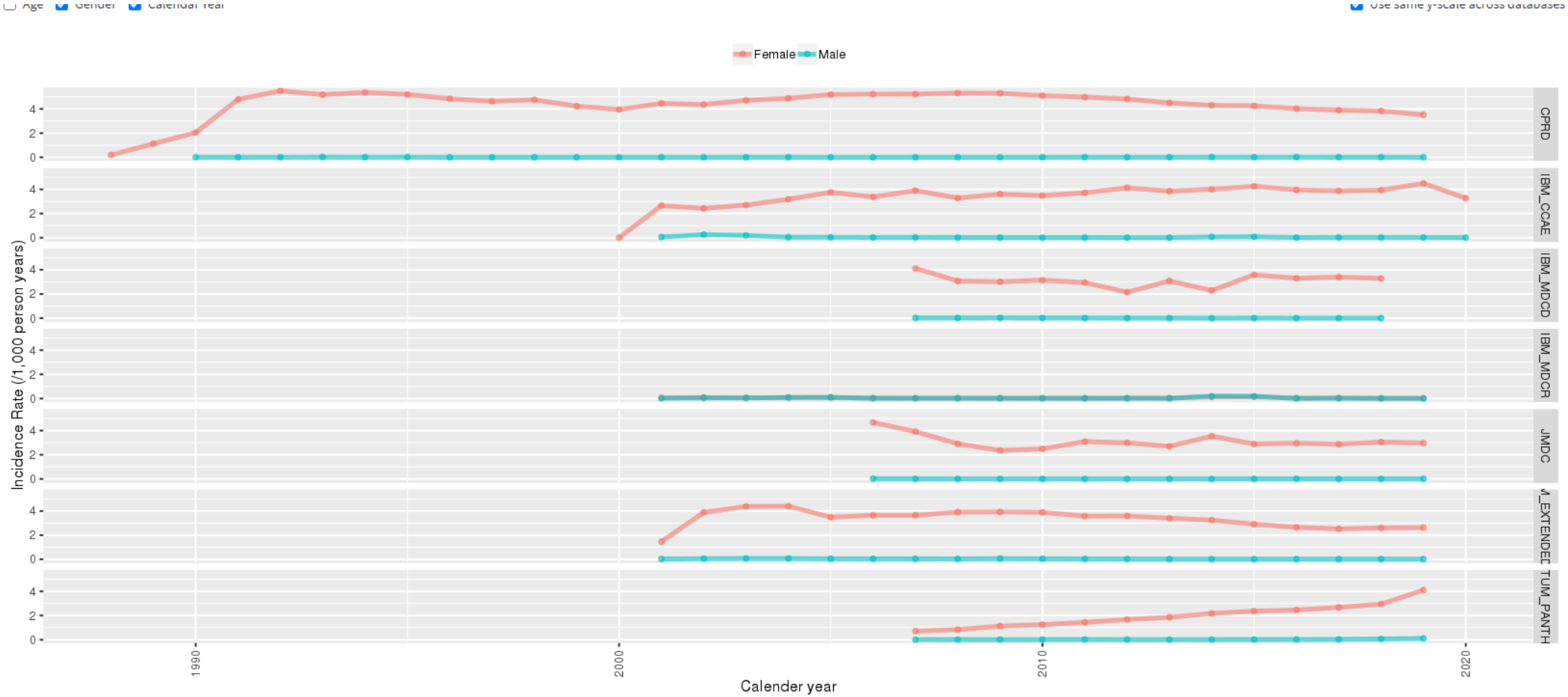
Functionality – Incidence Rate
Progressive increase over time



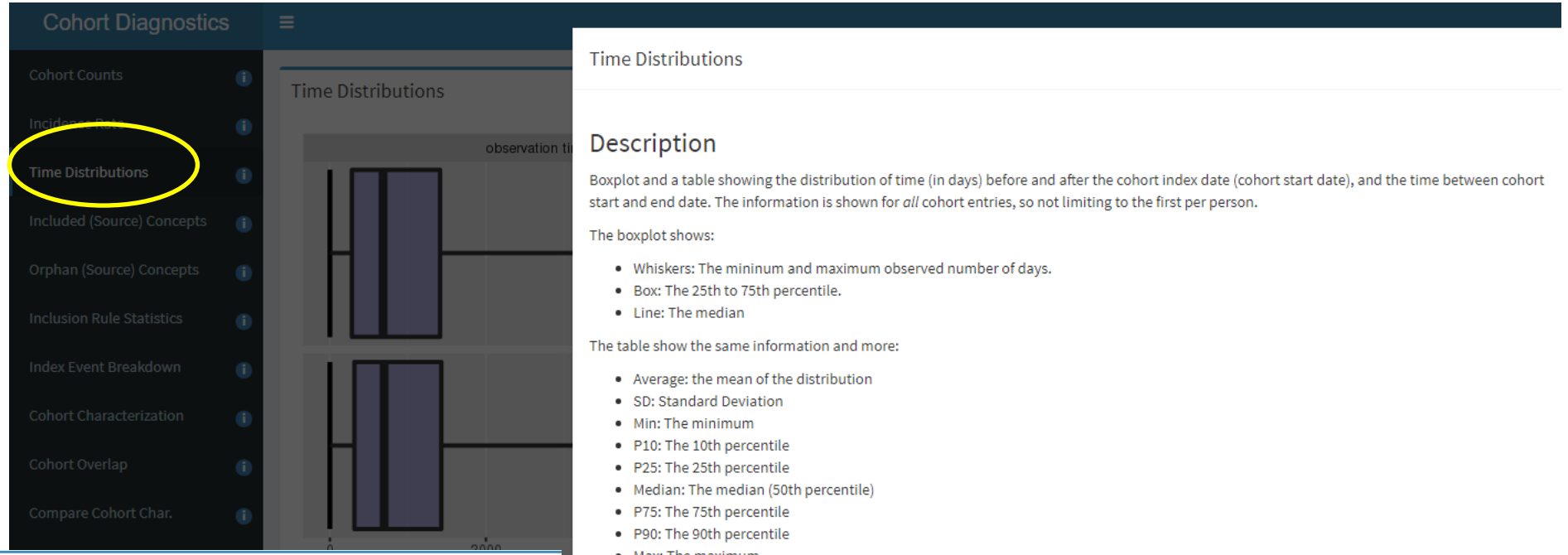
Functionality – Incidence Rate
Progressive decrease over time



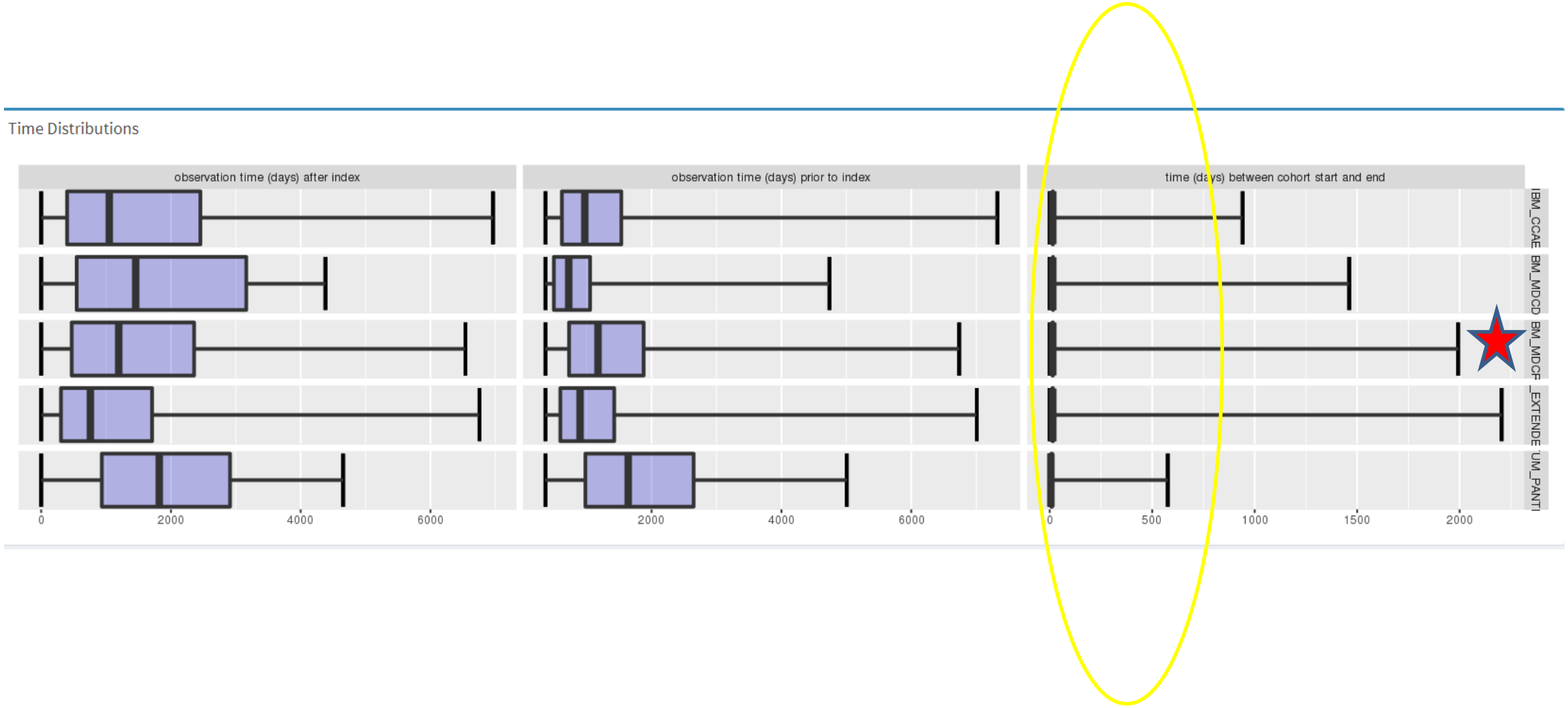
Functionality – Incidence Rate
Abortion among Male



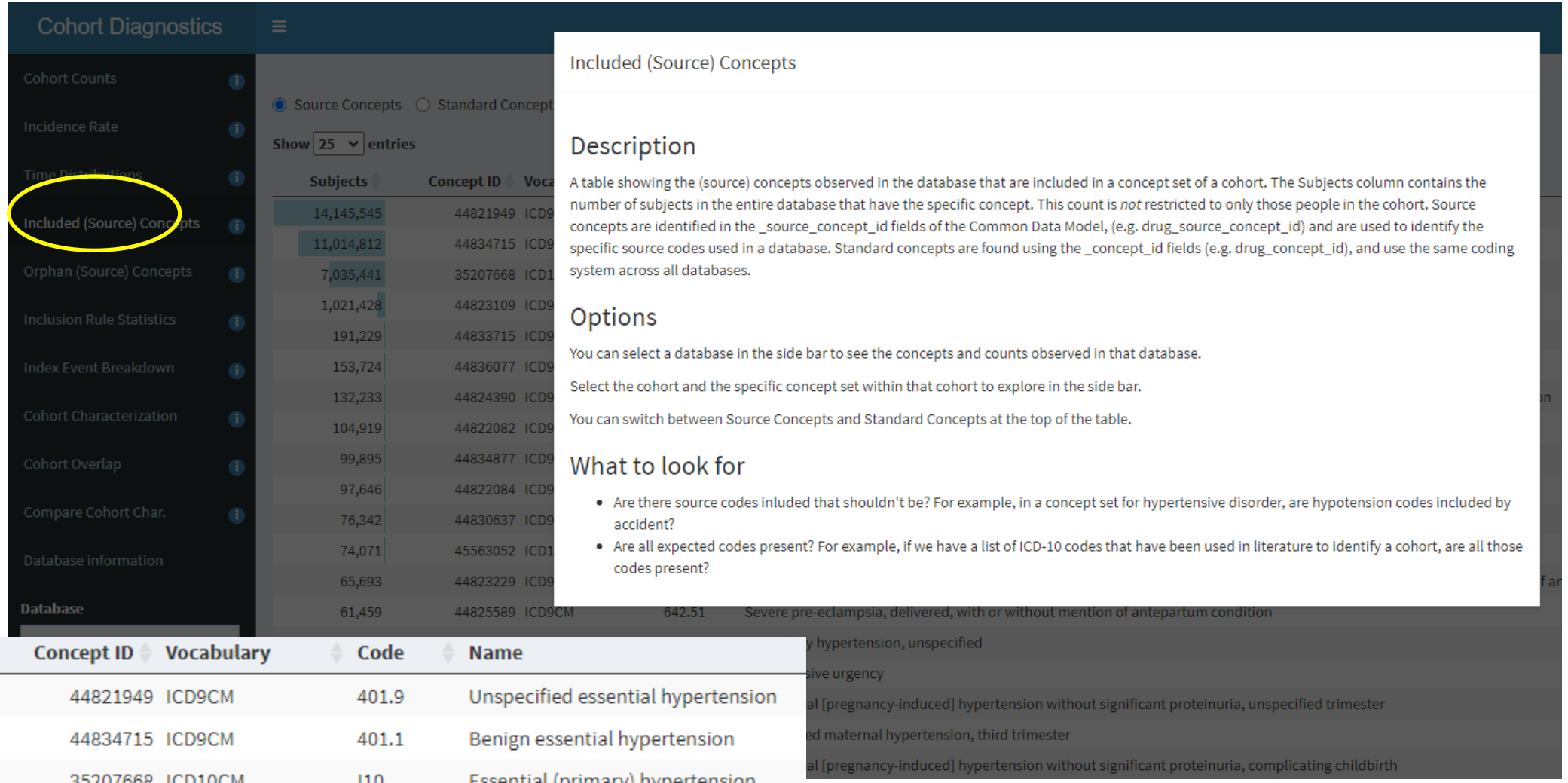
Functionality – Time distributions



Functionality – Time distributions

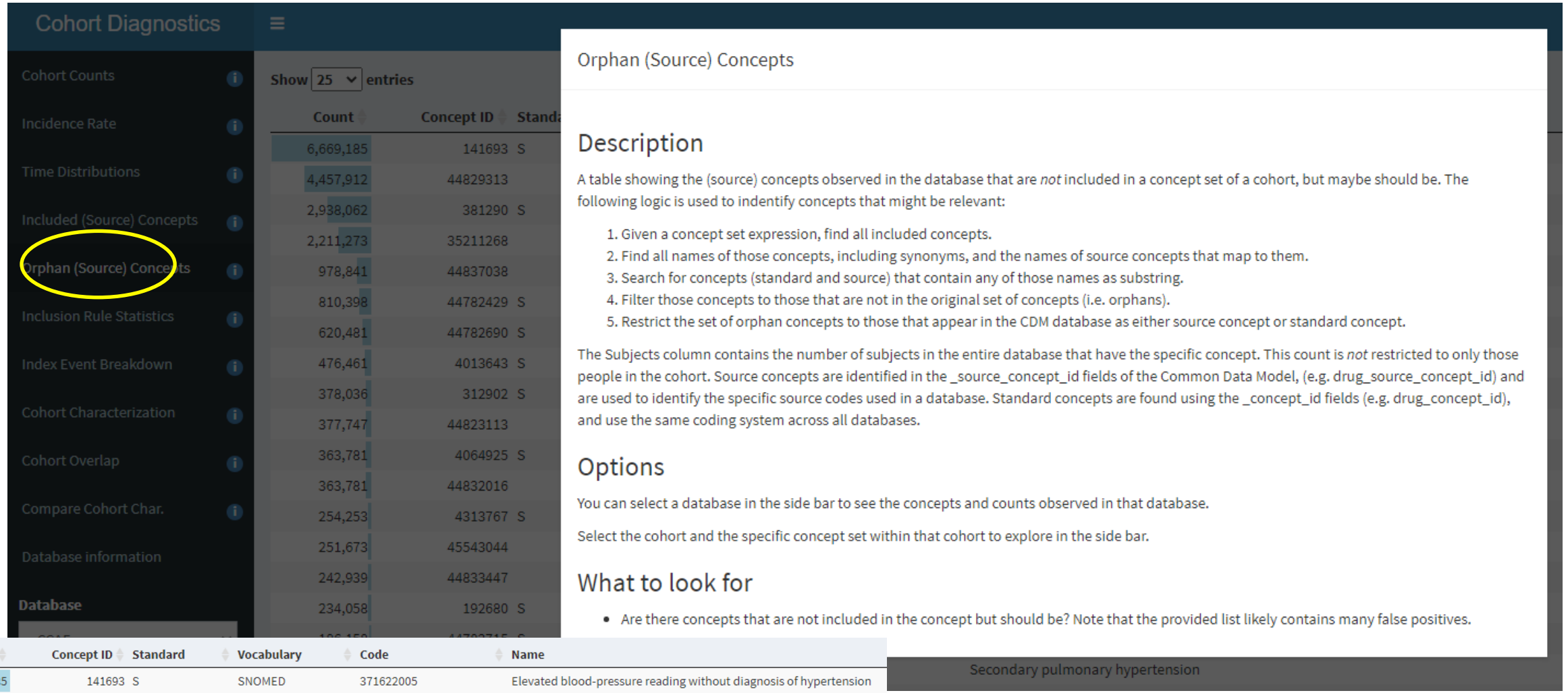


Functionality – Included concepts



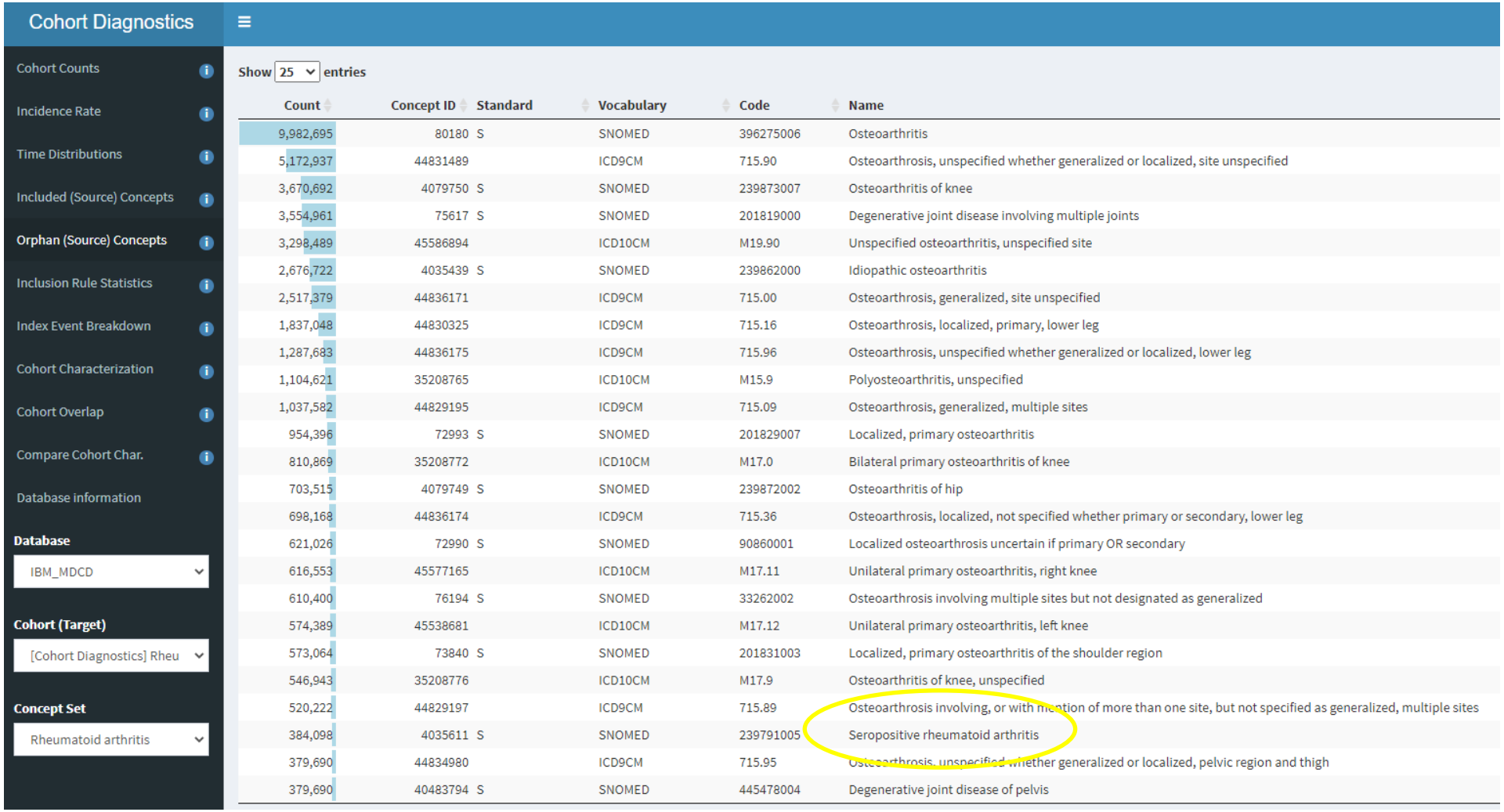
Subjects	Concept ID	Vocabulary	Code	Name
14,145,545	44821949	ICD9CM	401.9	Unspecified essential hypertension
11,014,812	44834715	ICD9CM	401.1	Benign essential hypertension
7,035,441	35207668	ICD10CM	I10	Essential (primary) hypertension
1,021,428	44823109	ICD9CM	401.0	Malignant essential hypertension

Functionality – Orphan concepts



Count	Concept ID	Standard	Vocabulary	Code	Name
6,669,185	141693	S	SNOMED	371622005	Elevated blood-pressure reading without diagnosis of hypertension
4,457,912	44829313		ICD9CM	796.2	Elevated blood pressure reading without diagnosis of hypertension
2,938,062	381290	S	SNOMED	4210003	Ocular hypertension
2,211,273	35211268		ICD10CM	R03.0	Elevated blood-pressure reading, without diagnosis of hypertension
978,841	44837038		ICD9CM	365.04	Ocular hypertension

Functionality – Orphan concepts



Count	Concept ID	Standard	Vocabulary	Code	Name
9,982,695	80180	S	SNOMED	396275006	Osteoarthritis
5,172,937	44831489		ICD9CM	715.90	Osteoarthrosis, unspecified whether generalized or localized, site unspecified
3,670,692	4079750	S	SNOMED	239873007	Osteoarthritis of knee
3,554,961	75617	S	SNOMED	201819000	Degenerative joint disease involving multiple joints
3,298,489	45586894		ICD10CM	M19.90	Unspecified osteoarthritis, unspecified site
2,676,722	4035439	S	SNOMED	239862000	Idiopathic osteoarthritis
2,517,379	44836171		ICD9CM	715.00	Osteoarthrosis, generalized, site unspecified
1,837,048	44830325		ICD9CM	715.16	Osteoarthrosis, localized, primary, lower leg
1,287,683	44836175		ICD9CM	715.96	Osteoarthrosis, unspecified whether generalized or localized, lower leg
1,104,621	35208765		ICD10CM	M15.9	Polyosteoarthritis, unspecified
1,037,582	44829195		ICD9CM	715.09	Osteoarthrosis, generalized, multiple sites
954,396	72993	S	SNOMED	201829007	Localized, primary osteoarthritis
810,869	35208772		ICD10CM	M17.0	Bilateral primary osteoarthritis of knee
703,515	4079749	S	SNOMED	239872002	Osteoarthritis of hip
698,168	44836174		ICD9CM	715.36	Osteoarthrosis, localized, not specified whether primary or secondary, lower leg
621,026	72990	S	SNOMED	90860001	Localized osteoarthrosis uncertain if primary OR secondary
616,553	45577165		ICD10CM	M17.11	Unilateral primary osteoarthritis, right knee
610,400	76194	S	SNOMED	33262002	Osteoarthrosis involving multiple sites but not designated as generalized
574,389	45538681		ICD10CM	M17.12	Unilateral primary osteoarthritis, left knee
573,064	73840	S	SNOMED	201831003	Localized, primary osteoarthritis of the shoulder region
546,943	35208776		ICD10CM	M17.9	Osteoarthritis of knee, unspecified
520,222	44829197		ICD9CM	715.89	Osteoarthrosis involving, or with mention of more than one site, but not specified as generalized, multiple sites
384,098	4035611	S	SNOMED	239791005	Seropositive rheumatoid arthritis
379,690	44834980		ICD9CM	715.95	Osteoarthrosis, unspecified whether generalized or localized, pelvic region and thigh
379,690	40483794	S	SNOMED	445478004	Degenerative joint disease of pelvis

Orphan codes

Concept set

Suicide attempt

Concept Name	Descendants
Suicide attempt	Y
Intentional harming self	Y

- Descendants
- Synonyms of descendants
- Source concepts mapping to included concepts

Search terms

Suicide attempt

Suicidal poisoning

Intentional self-harm

Self-harm

Self-inflicted injury

Orphans: matches not included in concept set

Electrocution, intent to self-farm

Search matches

Suicide attempt

Electrocution, intent to self-farm

Substring search

Concepts to search

Cough

Suicide attempt

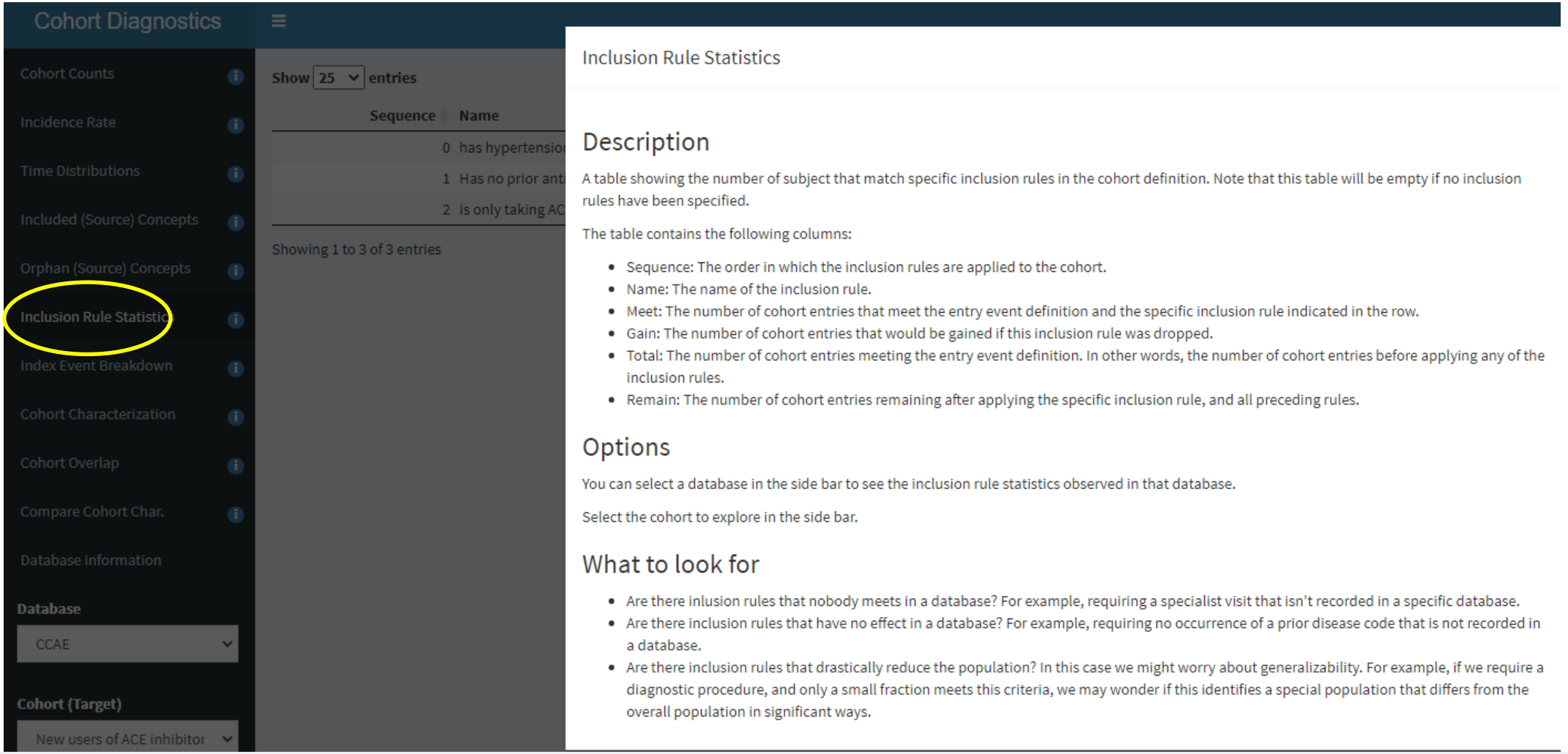
Acetaminophen

Electrocution, intent to self-farm

Data in Common Data Model

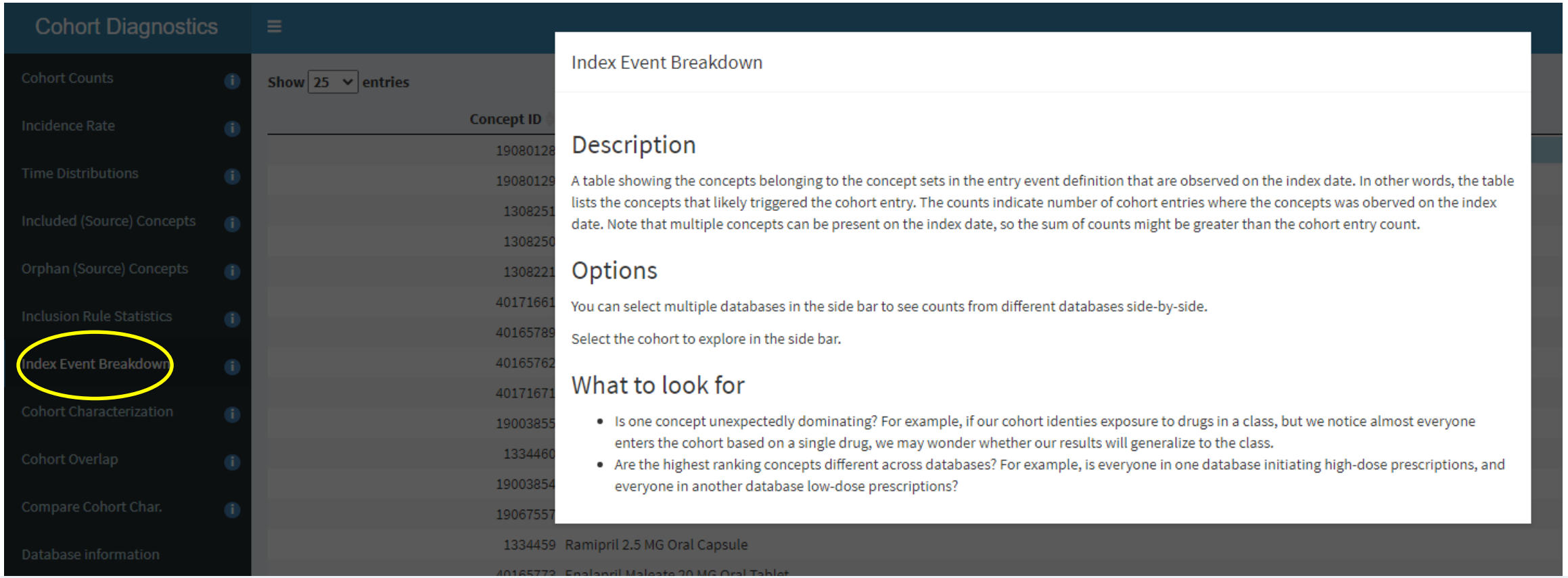
- All standard concepts
- All source concepts

Functionality – Inclusion Rule Statistics



Sequence	Name	Meet	Gain	Total	Remain
0	has hypertension diagnosis in 1 yr prior to treatment	2,547,647	539,522	3,604,155	2,547,647
1	Has no prior antihypertensive drug exposures in medical history	1,967,168	929,300	3,604,155	1,241,966
2	Is only taking ACE as monotherapy, with no concomitant combination treatments	2,555,382	395,771	3,604,155	846,195

Functionality – Index Event Breakdown



Concept ID	Name	CCAE Count	MDCD Count
19080128	Lisinopril 10 MG Oral Tablet	385,468	37,426
19080129	Lisinopril 20 MG Oral Tablet	154,099	16,531
1308251	Lisinopril 5 MG Oral Tablet	127,815	16,356
1308250	Lisinopril 2.5 MG Oral Tablet	23,487	4,189
1308221	Lisinopril 40 MG Oral Tablet	16,166	2,678

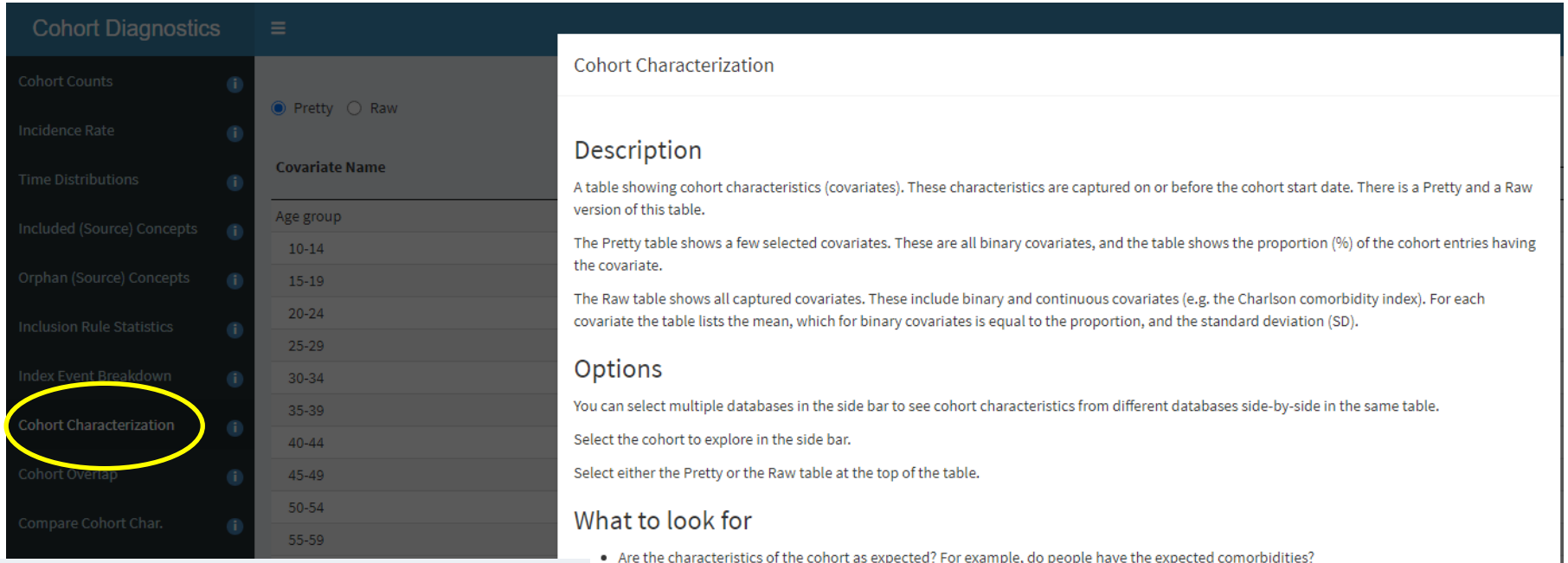
Functionality – Index Event Breakdown

Show 25 entries

Search:

Concept ID	Name	CPRD Count	IBM_CCAE Count	IBM_MDCD Count	IBM_MDCR Count	JMDC Count	OPTUM_EXTENDED_DOD Count	OPTUM_PANTHER Count
80809	Rheumatoid arthritis	76,744	446,615	108,530	135,411	116,722	650,465	997,312
80809	Rheumatoid arthritis	76,744	446,615	108,530	135,411	116,722	650,465	997,312
4083556	Seronegative rheumatoid arthritis	6,866	18,587	3,700	3,011	374	28,775	72,288
4083556	Seronegative rheumatoid arthritis	6,866	18,587	3,700	3,011	374	28,775	72,288
4035611	Seropositive rheumatoid arthritis	2,280	26,030	6,443	3,736	108	45,159	127,277
4035611	Seropositive rheumatoid arthritis	2,280	26,030	6,443	3,736	108	45,159	127,277
4114444	Flare of rheumatoid arthritis	710						111

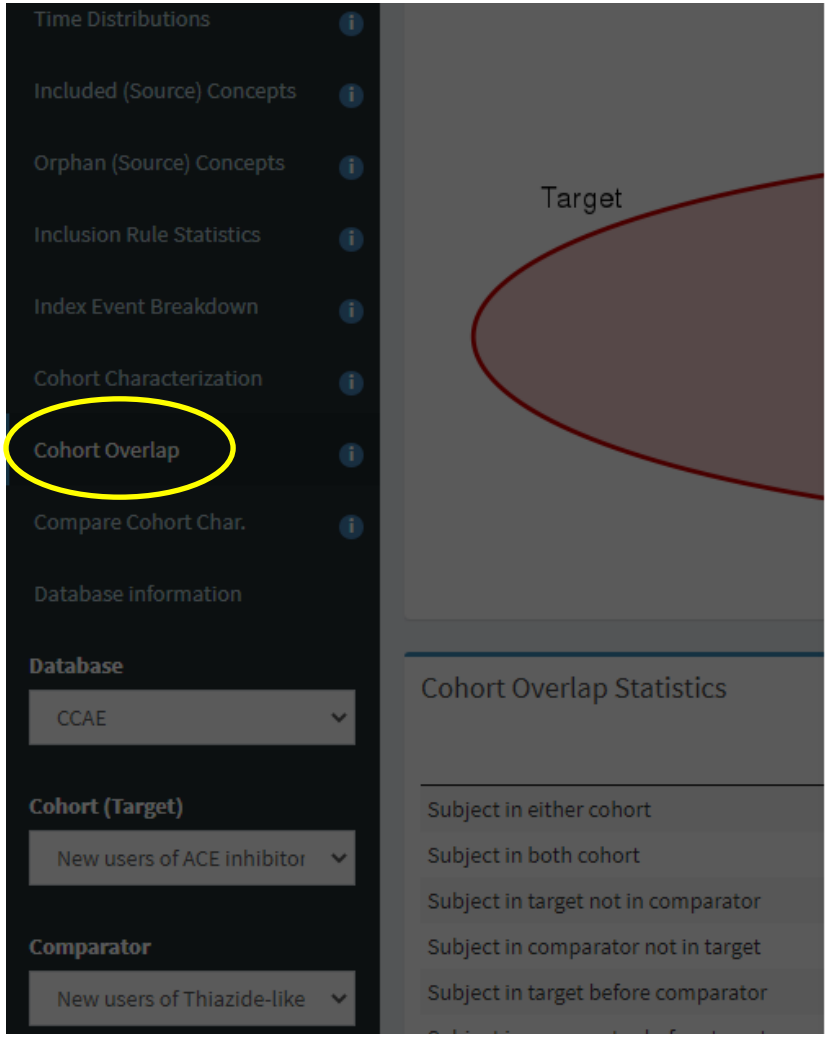
Functionality – Cohort Characterization



Pretty Raw

Covariate Name	CCAE
	Proportion
Age group	
10-14	0.2%
15-19	0.7%
20-24	1.4%
25-29	2.6%
30-34	5.0%
35-39	8.2%
40-44	12.0%
45-49	16.0%
50-54	18.6%
55-59	18.4%
60-64	15.5%
65-69	1.3%
Gender: female	38.3%
Medical history: General	
Acute respiratory disease	25.2%
Attention deficit hyperactivity disorder	1.2%
Chronic liver disease	1.4%
Chronic obstructive lung disease	1.6%
Crohn's disease	0.3%
Dementia	0.1%

Functionality - Cohort Overlap (Subjects)

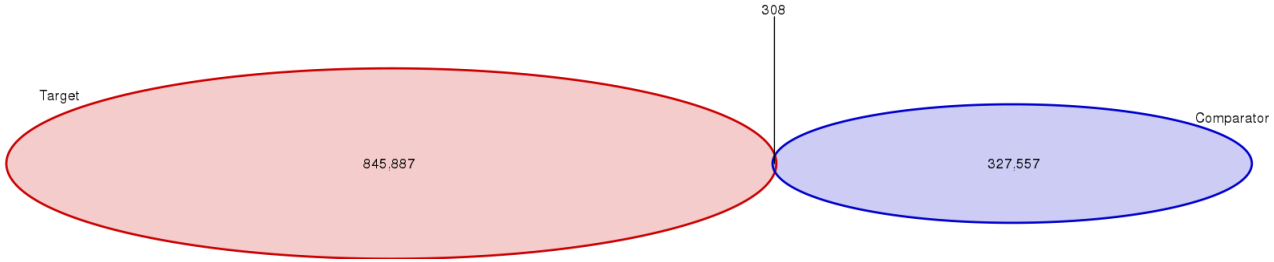


A Venn diagram showing the overlap between two cohorts, and a table listing several overlap statistics.

The Venn diagram shows the overlap in terms of subjects. It shows the number of subjects that belong to each cohort and to both. The diagram does *not* consider whether the subjects were in the different cohorts at the same time.

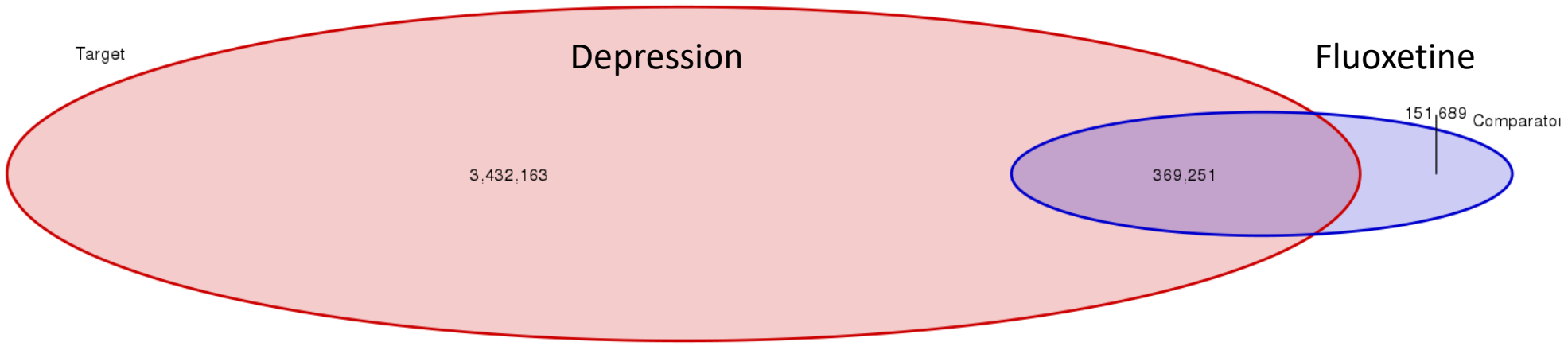
The table show the same information and more:

- Subject in either cohort: The number of subjects that enter one or both cohorts. (The union)
- Subject in both cohort: The number of subjects that enter both cohorts, although not necessarily at the same time. (The intersection)
- Subject in target not in comparator: The number of subjects that enter the target cohort, but not the comparator cohort. (Subtracting the comparator from the target)
- Subject in comparator not in target: The number of subjects that enter the comparator cohort, but not the target cohort. (Subtracting the comparator from the target)
- Subject in target before comparator: The number of subjects that enter both cohorts, but enter the target cohort before entering the comparator cohort. This number considers only the first entry per cohort per person.
- Subject in comparator before target: The number of subjects that enter both cohorts, but enter the comparator cohort before entering the target cohort. This number considers only the first entry per cohort per person.
- Subject in target and comparator on same day: The number of subjects that enter both cohorts on the same date. This number considers only the first entry per cohort per person.
- Subject having target start during comparator: The number of subjects that enter the target cohort during the comparator cohort, meaning comparator cohort start date <= target cohort start date <= comparator cohort end date. This number considers only the first entry per cohort per person.
- Subject having comparator start during target: The number of subjects that enter the comparator cohort during the target cohort, meaning target cohort start date <= comparator cohort start date <= target cohort end date. This number considers only the first entry per cohort per



Functionality - Cohort Overlap (Subjects)

Cohort Overlap (Subjects)

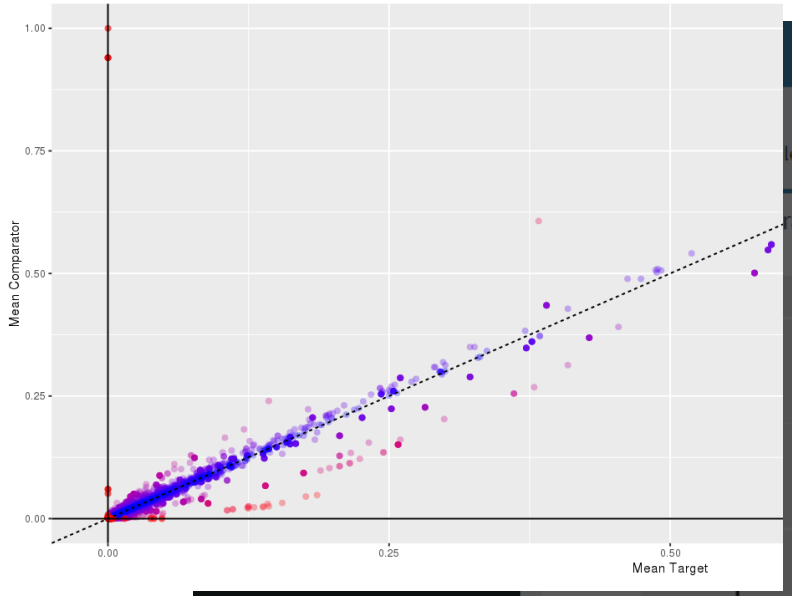


Cohort Overlap Statistics

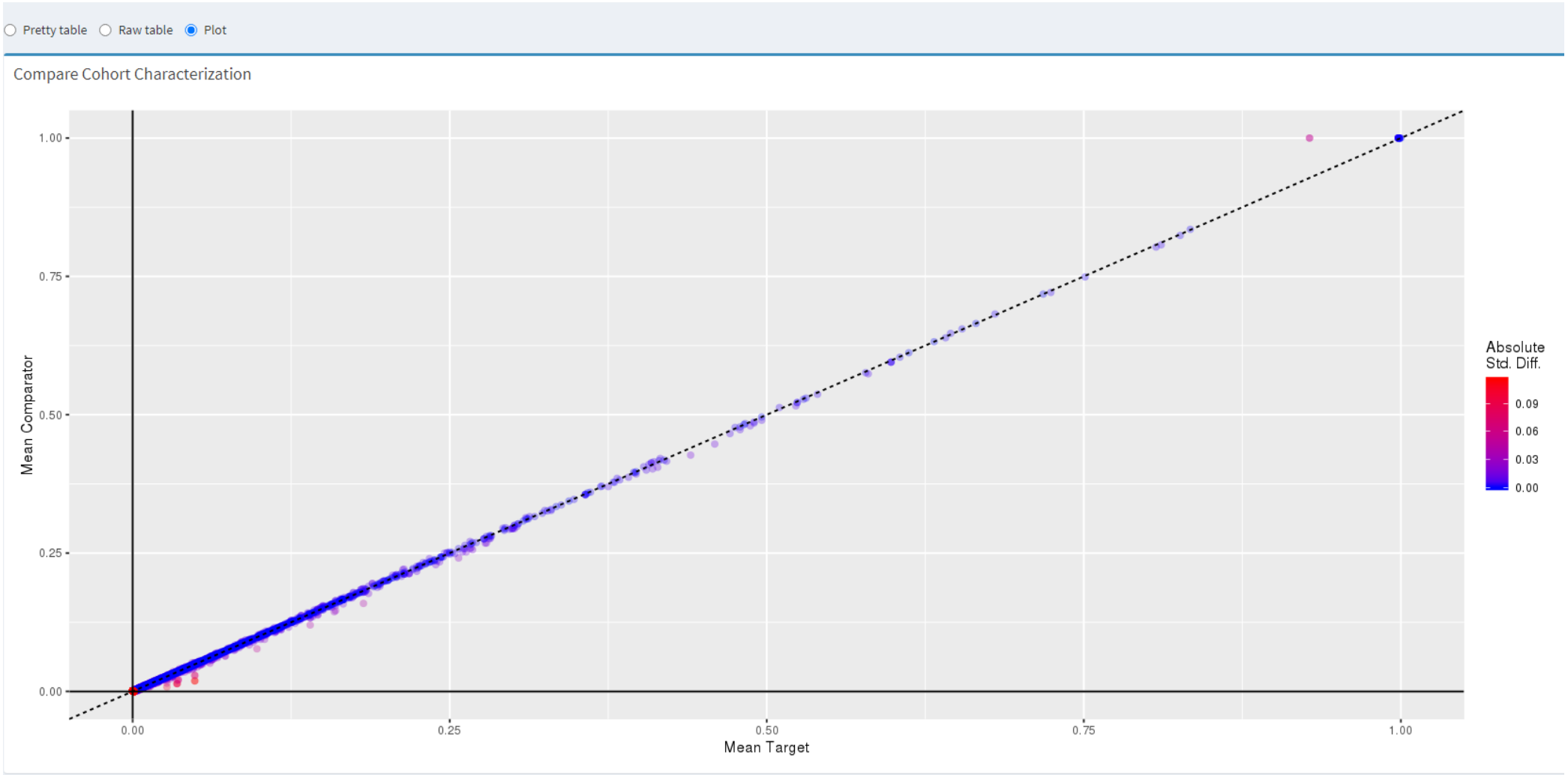
	Value
Subject in either cohort	3,953,103
Subject in both cohort	369,251
Subject in target not in comparator	3,432,163
Subject in comparator not in target	151,689
Subject in target before comparator	281,314
Subject in comparator before target	59,572
Subject in target and comparator on same day	28,365
Subject having target start during comparator	50,555
Subject having comparator start during target	280,654

Functionality – Compare Cohort Characteristics

Compare Cohort Characterization

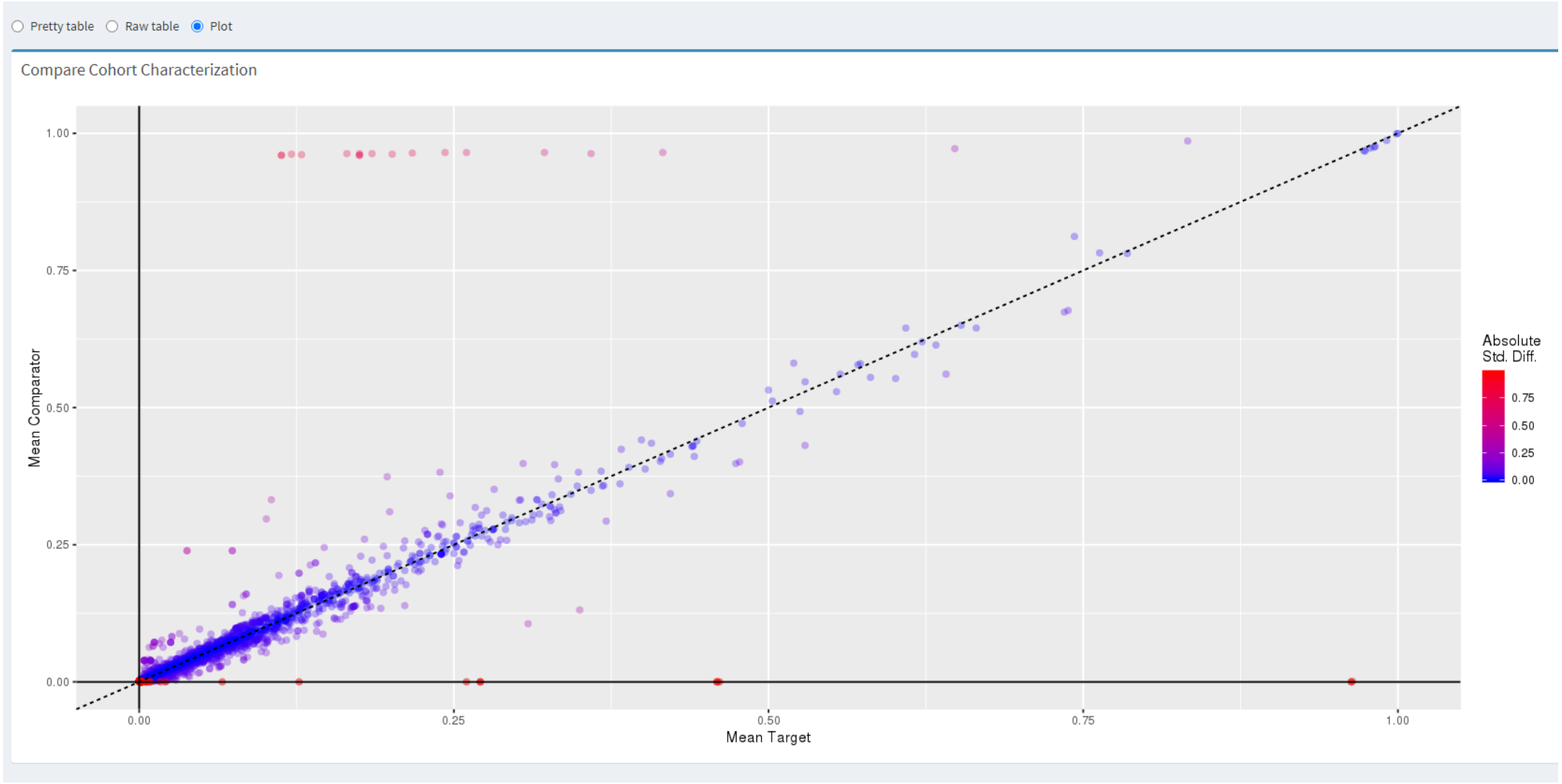


Functionality – Compare Cohort Characteristics



Functionality – Compare Cohort Characteristics

Ulcerative colitis

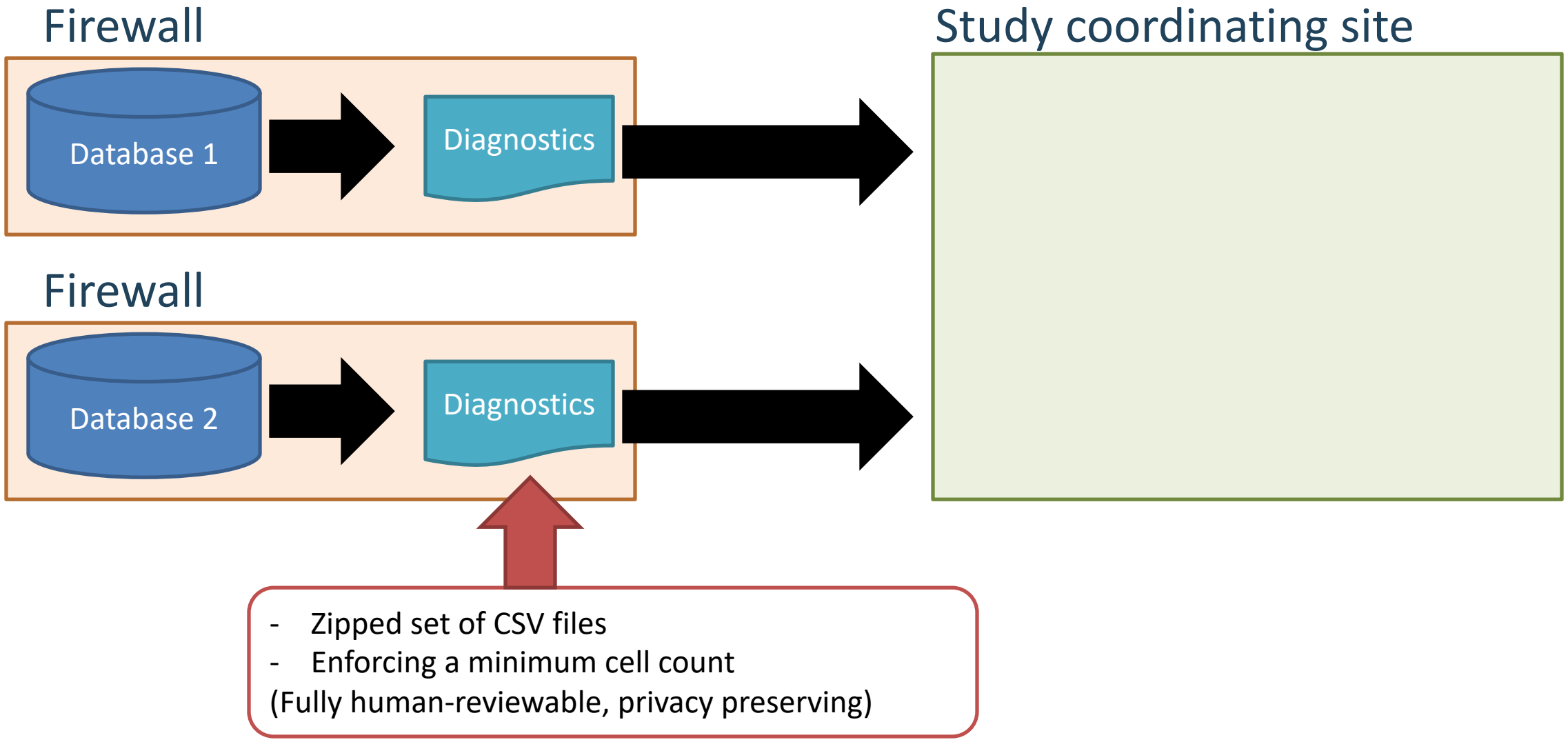


Crohns

Benefits of Cohort Diagnostics

- CohortDiagnostics has already proven instrumental in improving cohort definitions (how did we do without?)
- Provides insight into behavior of cohort definitions across research networks
- How to record and communicate the higher standard we're now using?

CohortDiagnostics in a distributed setting



Cohort Diagnostics

https://data.ohdsi.org/Covid19CohortEvaluationSafetyOutcomes/

Overview

- Motivation

- Uses cases → software functionality

Walk thru Use case

Backup

Clinical review

Gell and Coombs classification of immunologic drug reactions

Type	Description	Mechanism	Clinical features
I Immediate reaction (within one hour)	IgE-mediated, immediate-type hypersensitivity	Antigen exposure causes IgE-mediated activation of mast cells and basophils, with release of vasoactive substances, such as histamine, prostaglandins, and leukotrienes.	Anaphylaxis Angioedema Bronchospasm (hives) ion
II			anemia cytopenia nia
III			ckness action
IV	Cell-mediated or delayed hypersensitivity	Antigen exposure activates T cells, which then mediate tissue injury. Depending upon the type of T cell activation and the other effector cells recruited, different subtypes can be differentiated (ie, types IVa to IVd).	Contact dermatitis Some morbilliform reactions Severe exfoliative dermatoses (eg, SJS/TEN) AGEP DRESS/DiHS Interstitial nephritis Drug-induced hepatitis Other presentations

IgE: immunoglobulin E; Fc IgG: Fc portion of immunoglobulin G; SJS/TEN: Stevens-Johnson syndrome/toxic epidermal necrolysis; AGEP: acute generalized exanthematous pustulosis; DRESS/DiHS: drug rash with eosinophilia and systemic symptoms/drug-induced hypersensitivity syndrome.

Adapted from: Weiss ME, Adkinson NF. Immediate hypersensitivity reactions to penicillin and related antibiotics. Clin Allergy 1988; 18:515.

Complex syndrome – mostly clinical diagnosis (no testing)

- Skin manifestations
- Joint manifestations
- Fever

Data review

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

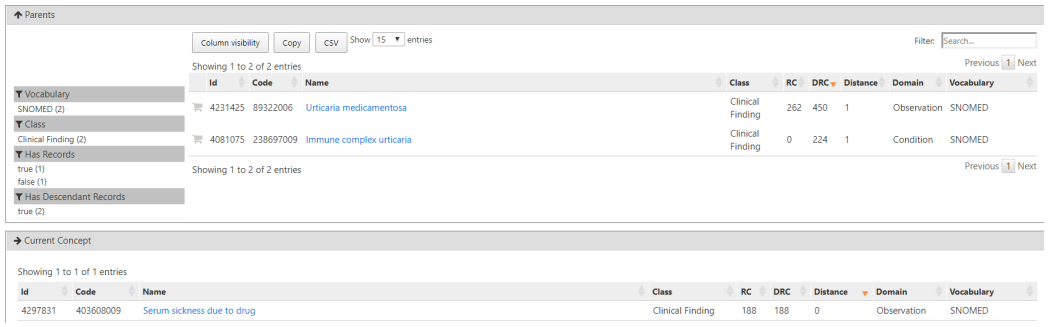
Showing 1 to 4 of 4 entries | Previous 1 Next

	Id	Code	Name	Class	RC	DRC	Domain	Vocabulary
	4297831	403608009	Serum sickness due to drug	Clinical Finding	188	188	Observation	SNOMED
	4297825	402658008	Serum sickness type vasculitis	Clinical Finding	36	36	Condition	SNOMED
	40410225	213322006	Serum sickness	Clinical Finding	0	0	Condition	SNOMED
	40410227	213324007	Serum sickness	Clinical Finding	0	0	Condition	SNOMED

Showing 1 to 4 of 4 entries | Previous 1 Next

Not coded in data!!

Hierarchy



Optum Extended SES (v1108) Observation Report

Prevalence

Treemap | Table

Column visibility | Copy | CSV | Show 15 entries

Showing 0 to 0 of 0 entries (filtered from 3,535 total entries)

Concept Id | Name

Showing 0 to 0 of 0 entries (filtered from 3,535 total entries)

IBM CCAE (v1103) Observation Report

Prevalence

Treemap | Table

Column visibility | Copy | CSV | Show 15 entries

Showing 0 to 0 of 0 entries (filtered from 3,229 total entries)

Optum PanTHER (v1109) Observation Report

Prevalence

Treemap | Table

Column visibility | Copy | CSV | Show 15 entries

Showing 1 to 1 of 1 entries (filtered from 6,543 total entries)

Concept Id | Name

4297831 Serum sickness due to drug

Showing 1 to 1 of 1 entries (filtered from 6,543 total entries)

Literature review

AMERICAN JOURNAL OF EPIDEMIOLOGY Vol. 132, No. 2

SERUM SICKNESS IN CHILDREN AFTER ANTIBIOTIC EXPOSURE: ESTIMATES OF OCCURRENCE AND MORBIDITY IN A HEALTH MAINTENANCE ORGANIZATION POPULATION

SUSAN R. HECKBERT,[1] W. SCOTT STRYKER,[1,3] KATHRYN L. COLTIN,[2] JOANN E. MANSON,[1,3] AND RICHARD PLATT[1]

Antibiotics	Rate per 1,000 person years	Proportion per 100 persons
No recent antibiotics	0.0001%	
Amoxicillin	0.0010%	0.01%
TMP-SMZ	0.0123%	0.09%
Cefaclor	0.0193%	0.14%
Penicillin V	0.0059%	0.04%

*20 days of starting antibiotic
Observational data
Case definition:
1) Urticaria, hives, rash (including erythema multiforme), or angioedema
2) arthritis, arthralgia or new join process

In 1990!

reported more frequently in children than in adults with an overall occurrence ranging from 0.5% (1 in 200) in one trial, to 0.024% (2 in 8,346) in overall clinical trials (with an incidence in children in clinical trials of 0.055%). The worldwide reporting rate for serum-sickness-like reactions in adults is very rare (<0.01%).

Product label

Drug	Rate per 1000 person years	Proportion per 100 persons
Amoxicillin		
TMP-SMZ		
Cefaclor		0.02%
Penicillin V		
Phenytoin		

Estimate from spontaneous report: 0.0018% (100x lower) (Platt et.al)

Very rare phenotype!

Lets start building phenotypes

- Broad definition vs Narrow definition
  - What is broad? What is narrow?
  - How about ‘Very broad’? Or even ‘Very **very** broad’
  - How about ‘very narrow’? Or ‘very **very** narrow’!

Since this is a syndrome – lets break it down into **‘Components’**

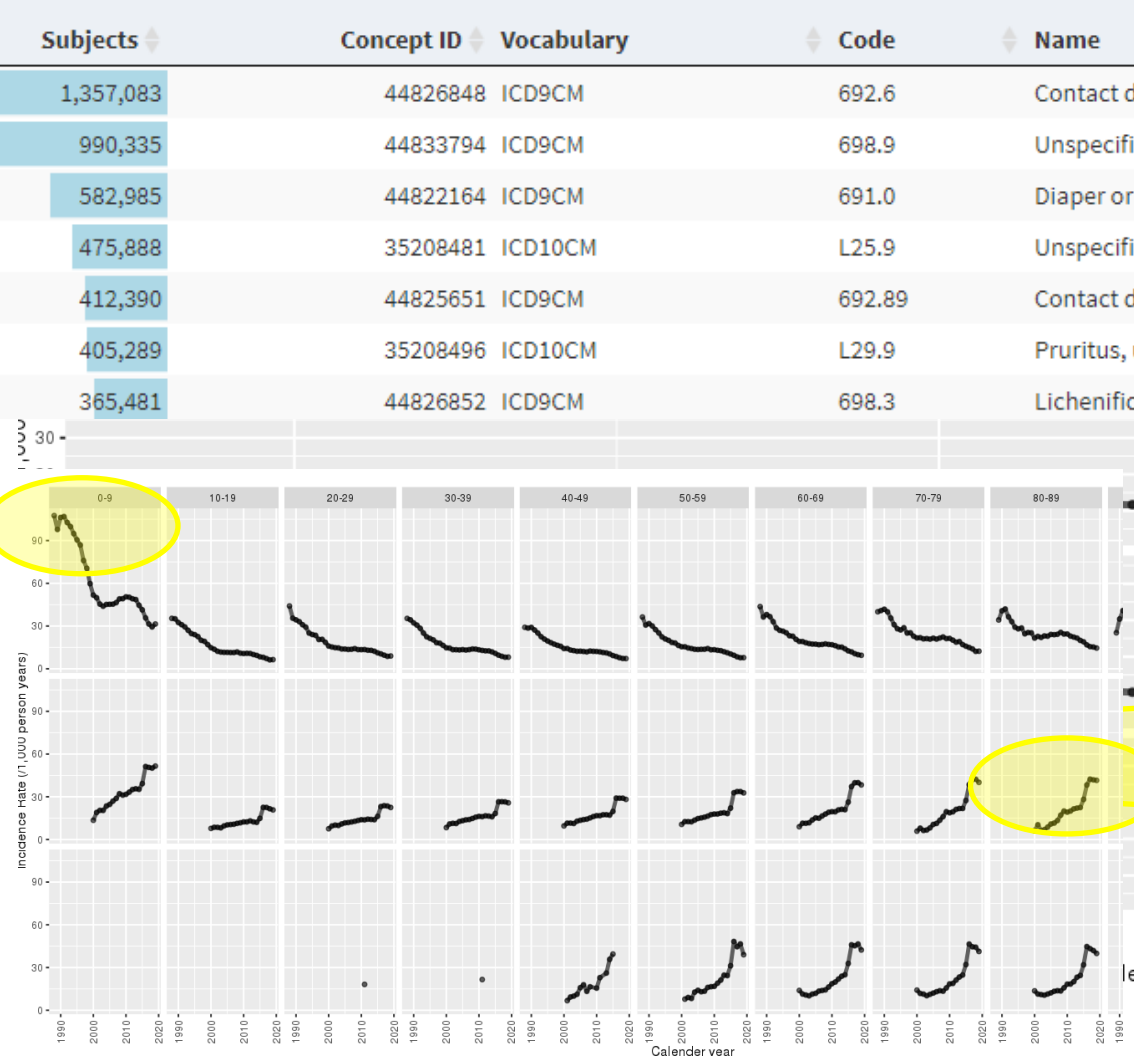
Syndrome – ‘combination of components temporally related’

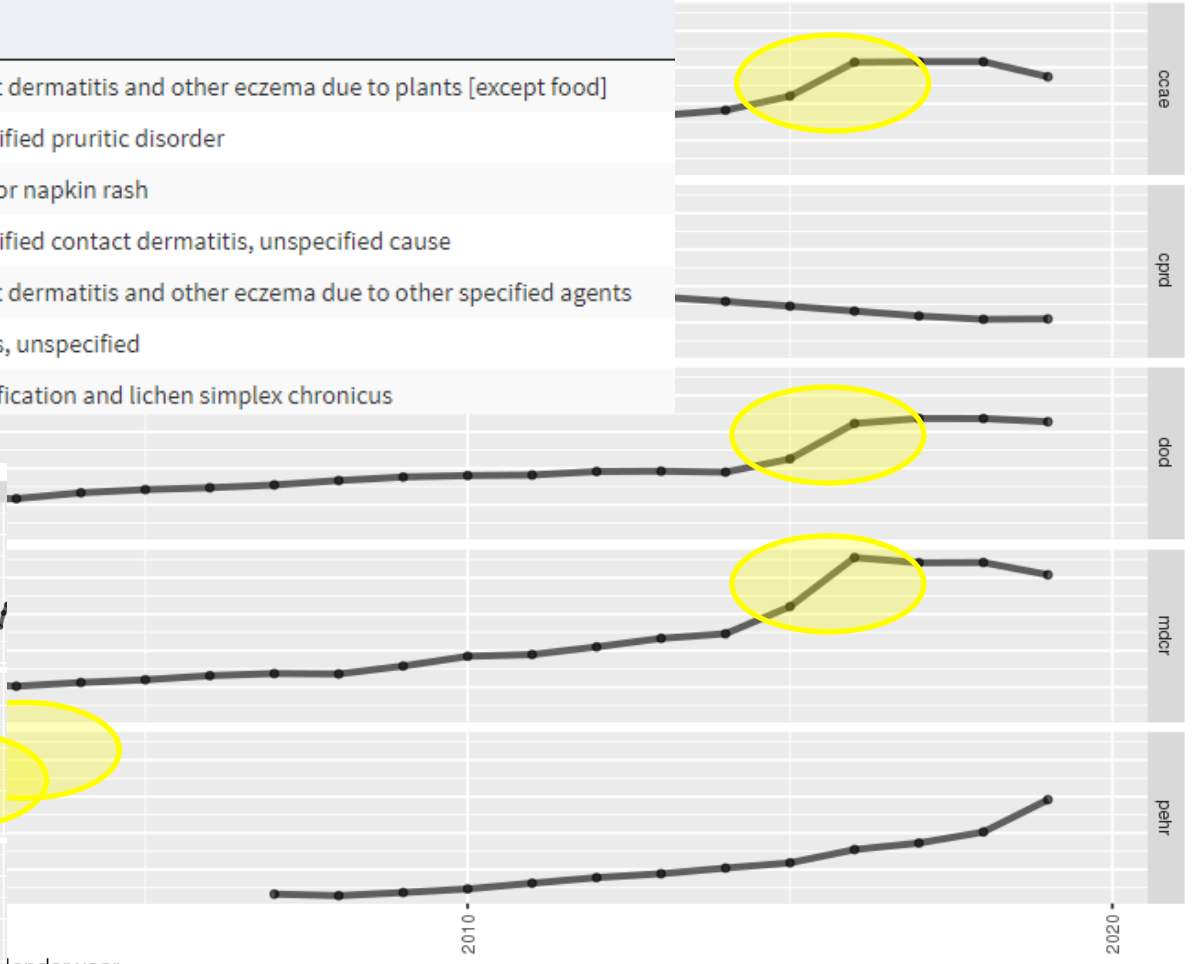
Symptoms	Signs	Clinical diagnosis	Common treatment/investigations
Itch/Pruritis	**Rash (2)**	Type 1 HS reaction	Anti-histamines (diphenhydramine, cetirizine)
Polyarthralgia (1)	**Fever (3) +/-**	Type 2 HS reaction	Steroids (Hydrocortisone, prednisone)
Joint pain	Hypotension	Type 3 HS reaction	Skin biopsy
Joint Swelling	Tachycardia	Type 4 HS reaction	Complement testing
Hives	Angioedema	**Polyarthritis (1)**	
Wheezing	Exanthema	Small joint arthritis	
Arthus reaction	Purpura	Large joint arthritis	
	Neutrophilia/Neutropenia Eosinophilia Thrombocytopenia	Anaphylaxis	
	Bronchospasm	Steven Johnson Syndrome	
	Urticaria (2)	Toxic Epidermal Necrolysis	
		Vasculitis	
		Photosensitivity	
		sweets syndrome	
		Contact dermatitis	
		Eczema	

Component - Itch

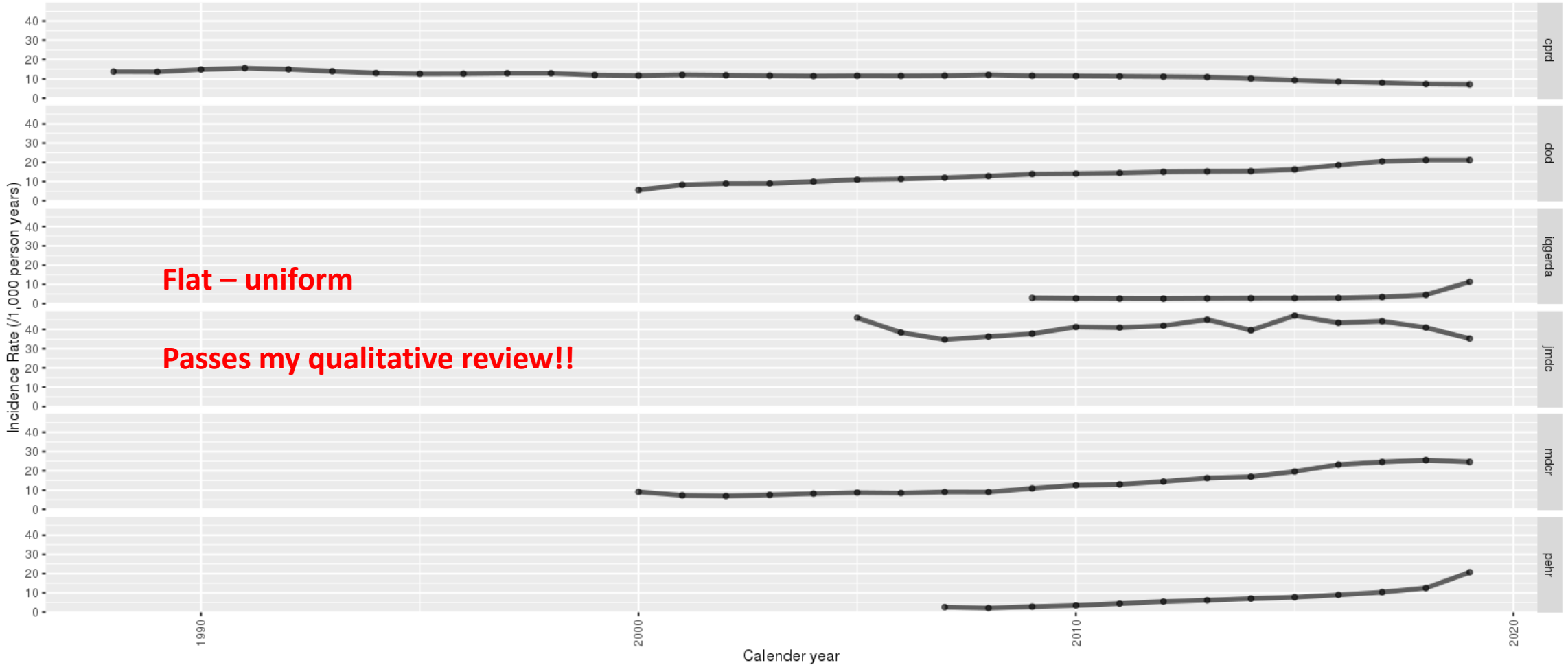
Show 25 entries

Subjects	Concept ID	Vocabulary	Code	Name
1,357,083	44826848	ICD9CM	692.6	Contact dermatitis and other eczema due to plants [except food]
990,335	44833794	ICD9CM	698.9	Unspecified pruritic disorder
582,985	44822164	ICD9CM	691.0	Diaper or napkin rash
475,888	35208481	ICD10CM	L25.9	Unspecified contact dermatitis, unspecified cause
412,390	44825651	ICD9CM	692.89	Contact dermatitis and other eczema due to other specified agents
405,289	35208496	ICD10CM	L29.9	Pruritus, unspecified
365,481	44826852	ICD9CM	698.3	Lichenification and lichen simplex chronicus





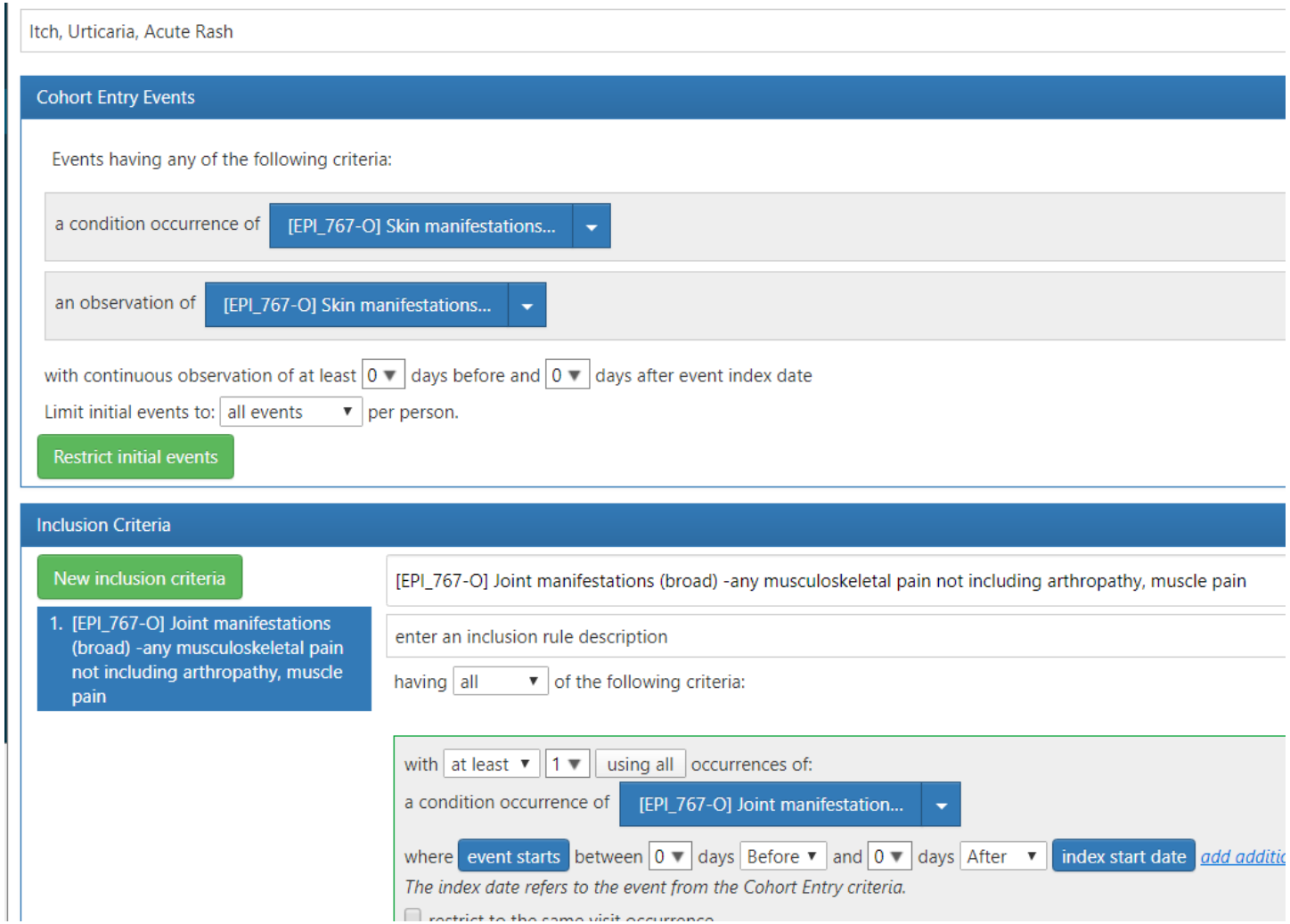
Itch or pruritis and-or urticaria (exclude diaper rash, contact dermatitis)



Rinse and repeat for each component

Symptoms	Signs	Clinical diagnosis	Common treatment/investigations
Itch/Pruritis	**Rash (2)**	Type 1 HS reaction	Anti-histamines (diphenhydramine, cetirizine)
Polyarthralgia (1)	**Fever (3) +/-**	Type 2 HS reaction	Steroids (Hydrocortisone, prednisone)
Joint pain	Hypotension	Type 3 HS reaction	Skin biopsy
Joint Swelling	Tachycardia	Type 4 HS reaction	Complement testing
Hives	Angioedema	**Polyarthritis (1)**	
Wheezing	Exanthema	Small joint arthritis	
Arthus reaction	Purpura	Large joint arthritis	
	Neutrophilia/Neutropenia Eosinophilia Thrombocytopenia	Anaphylaxis	
	Bronchospasm	Steven Johnson Syndrome	
	Urticaria (2)	Toxic Epidermal Necrolysis	
		Vasculitis	
		Photosensitivity	
		sweets syndrome	
		Contact dermatitis	
		Eczema	

Composite outcome definition



Composite definition