

Cohort Diagnostics

Gowtham Rao 06/16/2020

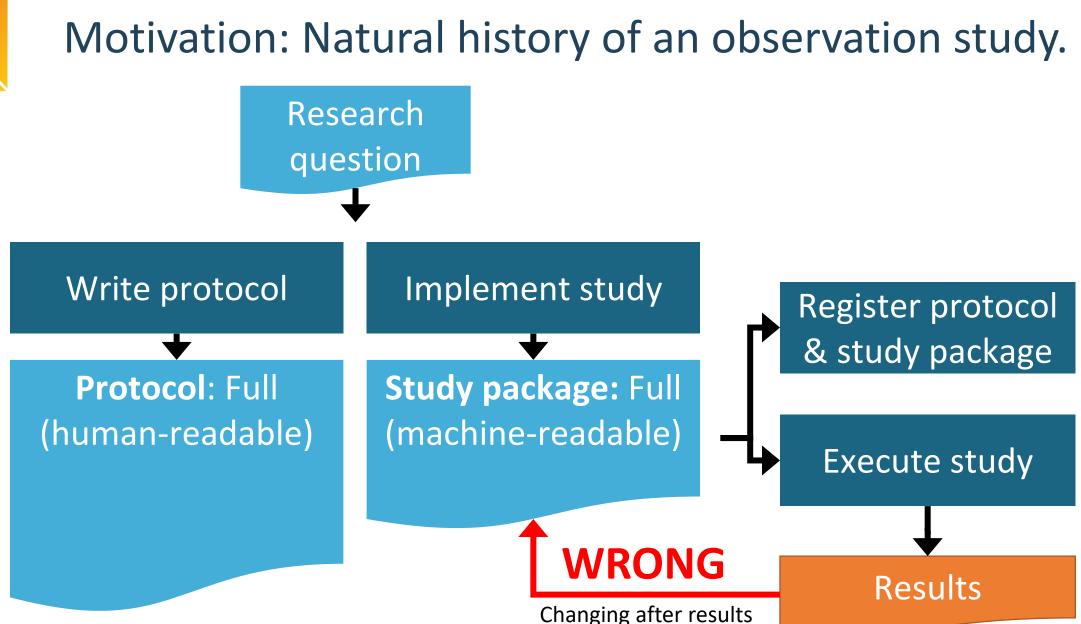


Overview

• Motivation

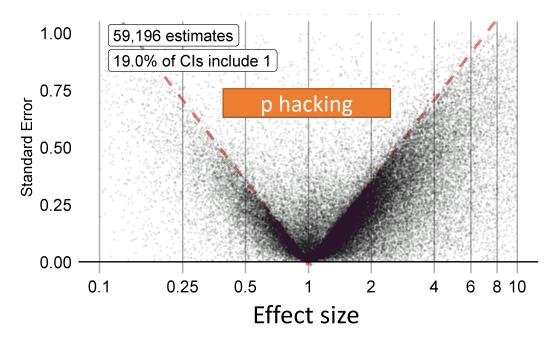
• Uses cases \rightarrow software functionality







Motivation: Reduce bias/Improve confidence



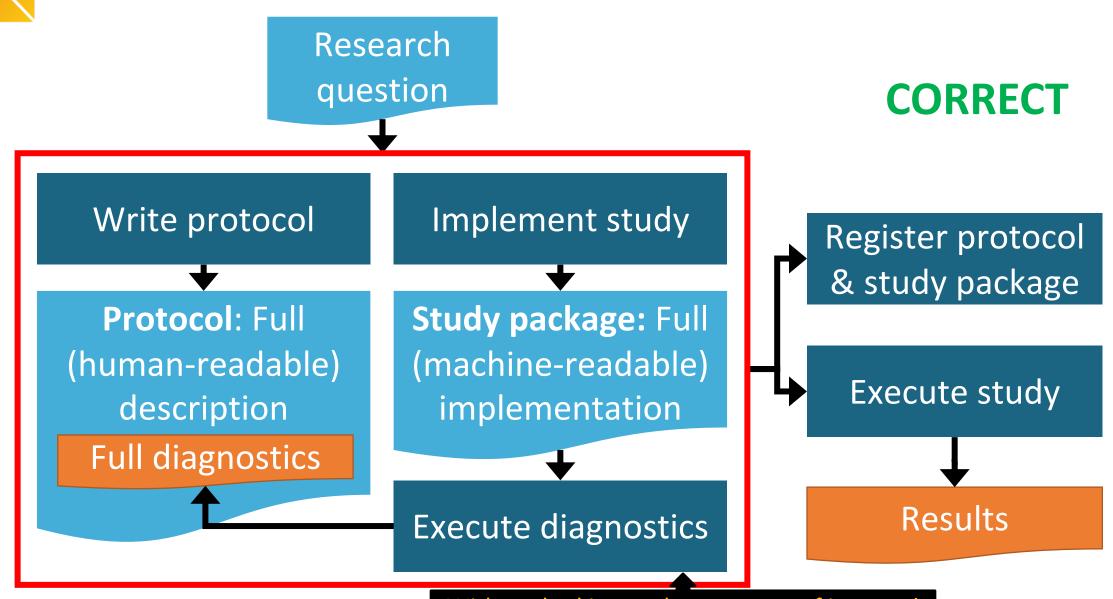
Effect size estimates in published observational studies

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



Without looking at the outcome of interest!

Diagnostics in OHDSI



Phenotype Diagnostics

- Clinical practice experience
- Data domain expertise
- Coding experience

PheValuator

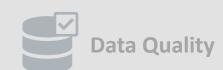
Something was missing

- 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.





Clinical Validity



Software Validity



Method Validity





• Motivation

Uses cases → software functionality



New R-package - Overview

Cohort Diagnostics		≡				
Cohort Counts	•	Show 25 🗸 entries				
Incidence Rate	•	Cohort	JMDC			
	-	· · · · · · · · · · · · · · · · · · ·	Entries 🔶	Subjects 🔷	Entries 🔶	Subjects 🔶
Time Distributions	•	New users of ACE inhibitors as first-line monotherapy for hypertension (2)	846195	846,195	2,691	2,691
Included (Source) Concepts	•	New users of Thiazide-like diuretics as first-line monotherapy for hypertension (2)	327865	327,865	778	778
		Angioedema outcome	136844	127,9 <mark>75</mark>	78	74
Orphan (Source) Concepts	•	Acute myocardial infarction outcome	516340	485,068	6,897	6,450
Inclusion Rule Statistics	•	ACEs with bad inclusion criterion				
inclusion rule statistics	•	Suicide attempt or ideation	571991	571,991	122	122
Index Event Breakdown	•	Showing 1 to 6 of 6 entries				
	-	Showing 1 to 6 of 6 entries				
Cohort Characterization	•					
Cohort Overlap	•	OHDSI / CohortDiagnostics			O Unwatch	· ▼ 11
	_				U UIIIIdei	·
Compare Cohort Char.	•					
Database information		<>Code (!) Issues 22 \$\$ Pull requests 1 (▷) Actions (!) Se	curity 0	Insights		
Database						
CCAE		An R package for performing various cohort diagnostics. https://d	obdsi aithub iq)/CohortDia	anos	
JMDC		An to package for performing various conort alagnostics. https://t	Shashgithabh	, conorcon	ignos	
MDCD		hades				
OHDSI Concept Prevalence						
✓ Synpuf						



Functionality – Cohort Counts

Cohort Diagnostics		=								
ohort Counts		Show 25 🗸 entries	Cohort Counts							
ncidence Rate		Cohort	Description		je					
		New users of ACE inhibitors as first-line mono		table showing the number of cohort entries and unique persons per cohort per database. Because one person can have more than one cohort entry,						
		New users of Thiazide-like diuretics as first-li	the number of entries can be higher than the number of persons.							
		Angioedema outcome	Options							
		Acute myocardial infarction outcome	You can select multiple databases in the side bar to see counts from different databases side-by-side.							
		ACEs with bad inclusion criterion Suicide attempt or ideation	What to look for							
ndex Event Breakdown ohort Characterization ohort Overlap	•	Showing 1 to 6 of 6 entries	 Are there cohorts that are empty in some databases? Are the relative counts (relative to the other cohorts in the same database) comparable across databases? Note that the color bars show the relative counts. Are the cohorts of expected and sufficient size? For example, if we want to study the effect of an exposure, a rule-of-thumb is that we require at least 2,500 in the exposure cohort. 							
or Cohort			€ -	Entries 🔶	Subjects					
New users of ACE	inhi	bitors as first-line monotherapy fo	or hypertension (2)	846195	846,1					
New users of Thia	azide	like diuretics as first-line monoth	herapy for hypertension (2)	327865	327,8					
Angioedema outcome 136844 127,										
Acute myocardial infarction outcome 516340 485,0										
ACEs with bad inclusion criterion										
Suicide attempt o	or ide	eation		571991	571,9					



Functionality – Cohort Counts

Cohort	CPRD		IBM_CCAE		IBM_MDCD		IBM_MDCR		JMDC		OPTUM_EXTENDED_DOD		OPTUM_PANTHER	
Colloit	Entries 🔶	Subjects 🔷	Entries 🔶	Subjects 🔷	Entries 🔶	Subjects 🔶	Entries 🔶	Subjects 🔶	Entries 🔶	Subjects 🔷	Entries 🔶	Subjects 🔶	Entries 🔶	Subjects 🔶
[Cohort Diagnostics] Bacterial Pneumonia	26374	23 , 86 <mark>0</mark>	1,311,797	854,098	749,091	435,83 <mark>3</mark>	669 , 16 <mark>2</mark>	352,32 <mark>6</mark>	20 <mark>8,258</mark>	89 <mark>,700</mark>	1,935,043	910,27 <mark>6</mark>	1,487,799	66851 <mark>1</mark>
[Cohort Diagnostics] Hospitalizations with Pneumonia			746,581	625,029	747,000	514,45 <mark>6</mark>	956,50 <mark>0</mark>	706, <mark>649</mark>	54,885	46 , 55 <mark>0</mark>	1,632,905	1,211,93 <mark>4</mark>	910,717	73002 <mark>8</mark>
[Cohort Diagnostics] Pneumonia	220467	193,475	4,136,169	3,278,6 <mark>70</mark>	2,157, <mark>561</mark>	1,39 <mark>5,845</mark>	2,068 <mark>,702</mark>	1,2 <mark>06,277</mark>	441,396	317,061	4,369,39 <mark>8</mark>	3,003, <mark>301</mark>	4,654,758	315 <mark>8412</mark>
[Cohort Diagnostics] Viral Pneumonia	30272	30,27 <mark>2</mark>	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	29,68 <mark>9</mark>	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	3,559,23 <mark>5</mark>	3,144,5 <mark>56</mark>	1,710,875	1,338 <mark>,548</mark>	1,570, <mark>292</mark>	1,16 <mark>1,432</mark>	330,695	277,160	3,490,79 <mark>2</mark>	2,871, <mark>062</mark>	3,798,0 <mark>57</mark>	307 <mark>3883</mark>

Concept Name	
--------------	--

Pneumonia

Ornithosis with pneumonia

Idiopathic pneumonia syndrome

Fungal pneumonia

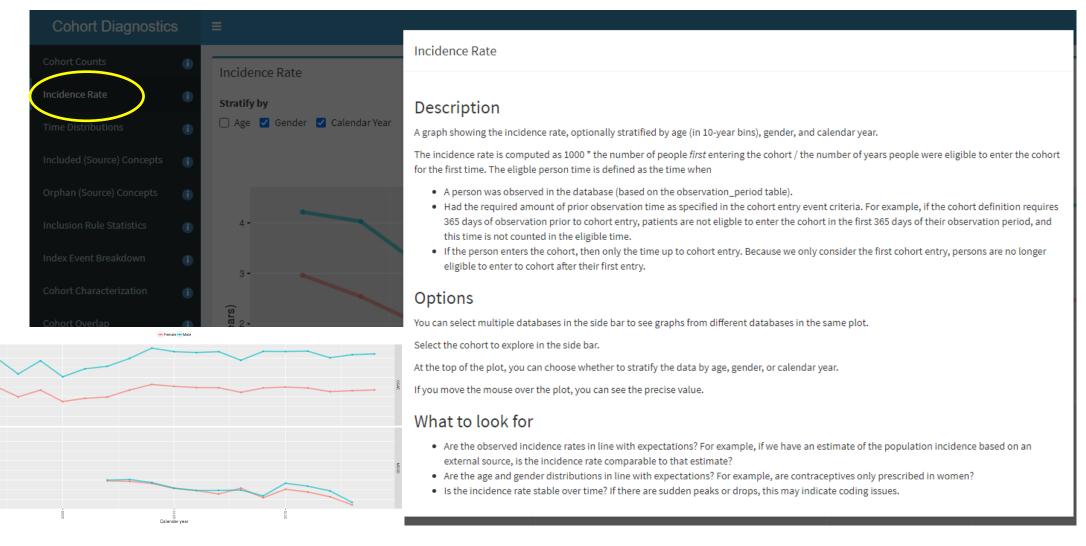
Atypical pneumonia

Acute ulcerative gastroenteritis complicating pneumonia

Acute pneumonia due to coccidioidomycosis

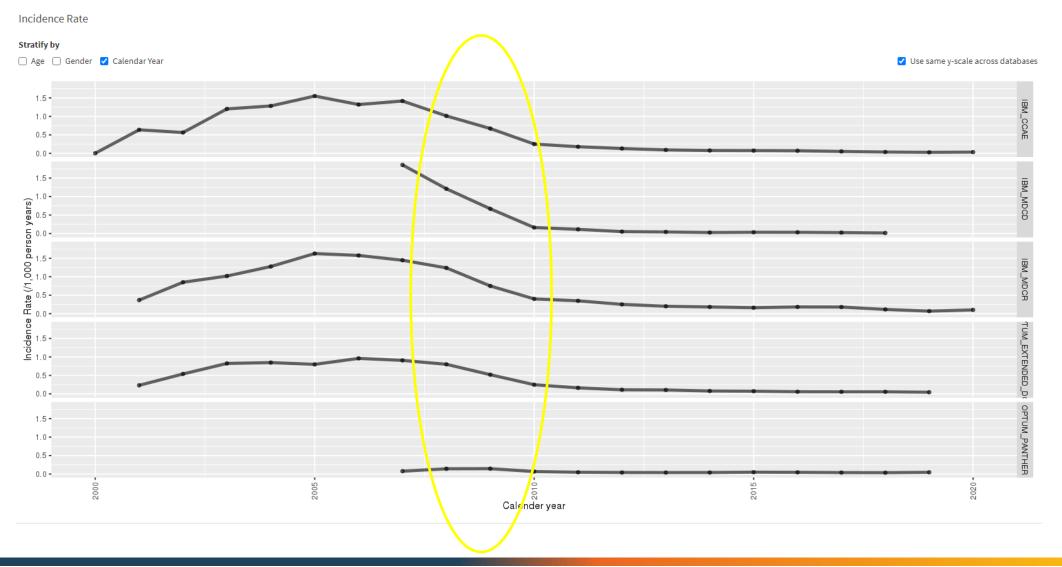


Functionality – Incidence Rate



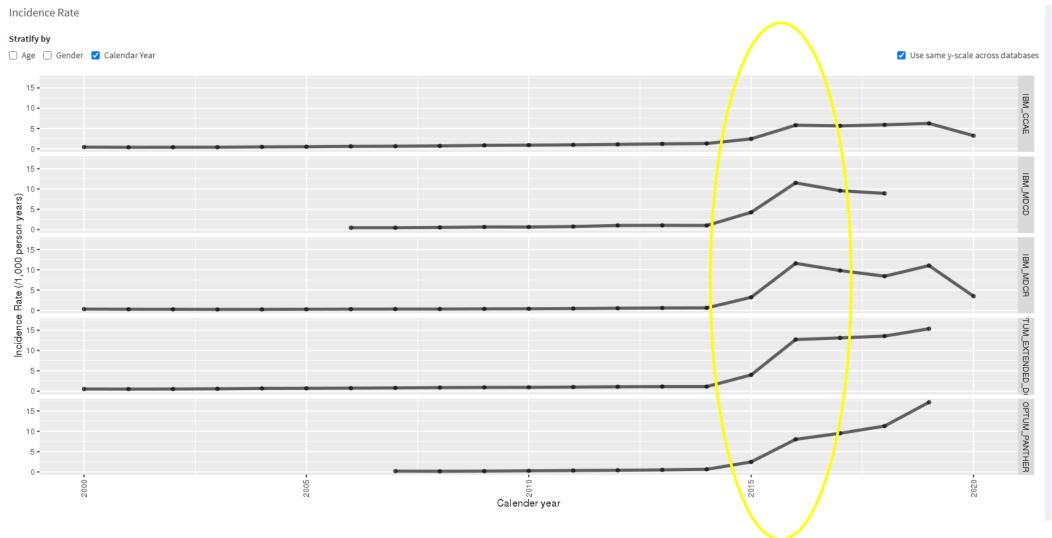


Functionality – Incidence Rate Dramatic drop around 2008



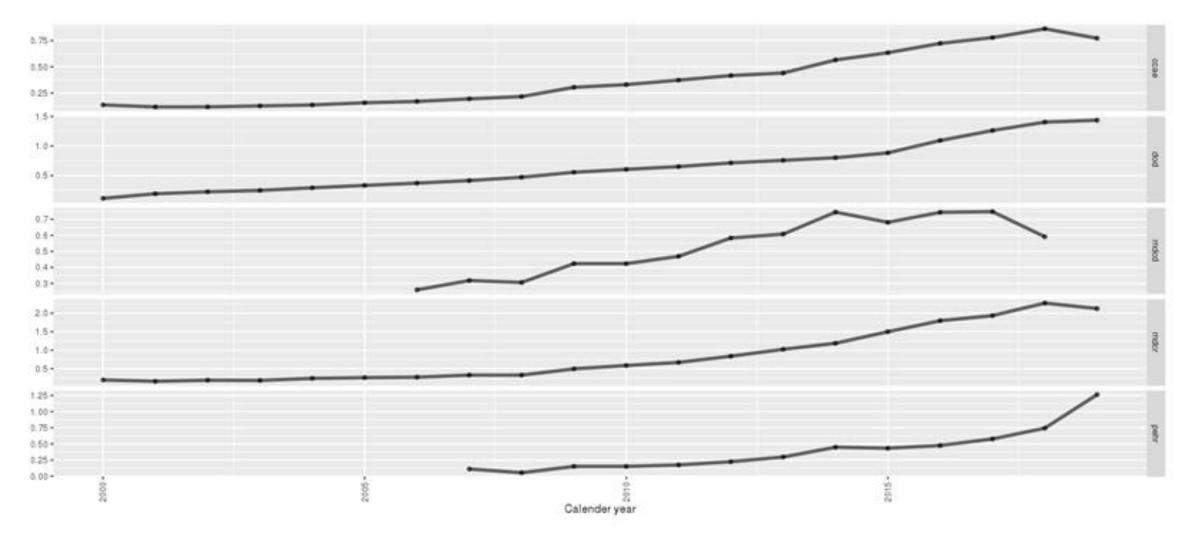


Functionality – Incidence Rate Increase around 2015



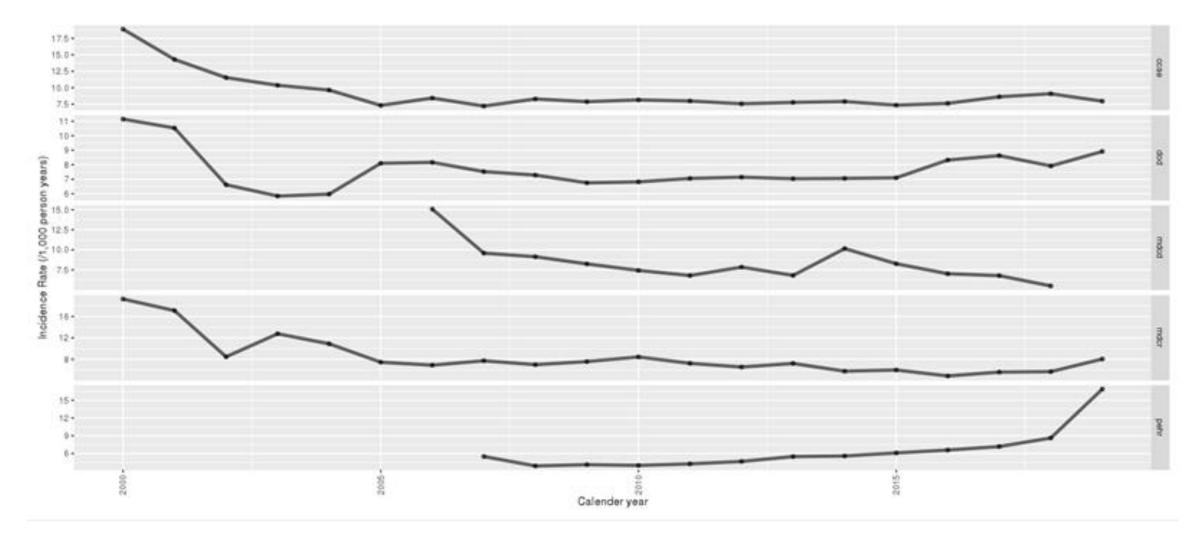


Functionality – Incidence Rate Progressive increase over time





Functionality – Incidence Rate Progressive decrease over time

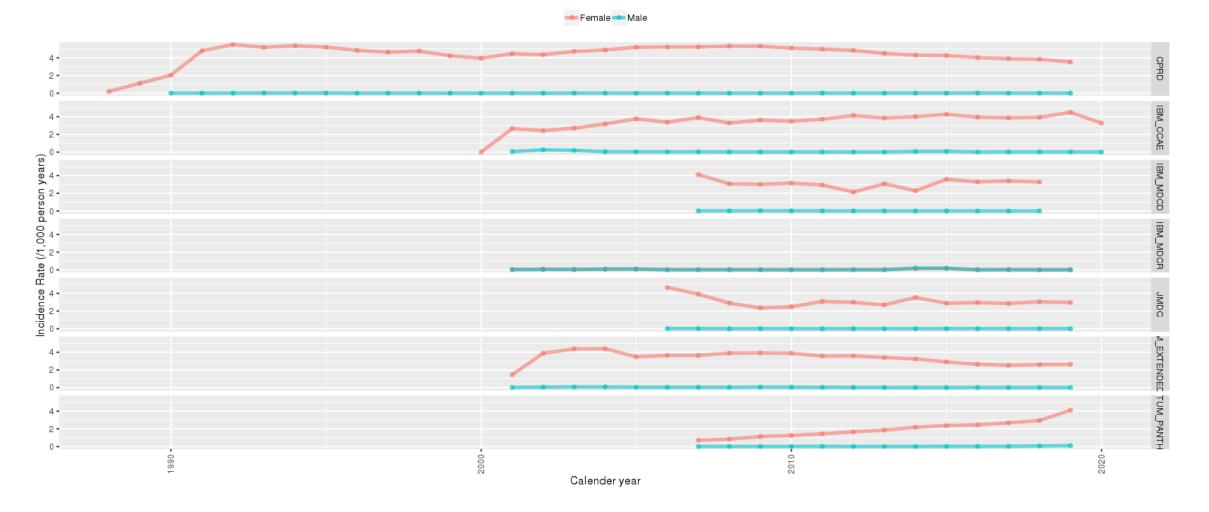




Functionality – Incidence Rate Abortion among Male

🗌 Age 🔽 Genuer 🔽 Catenual real

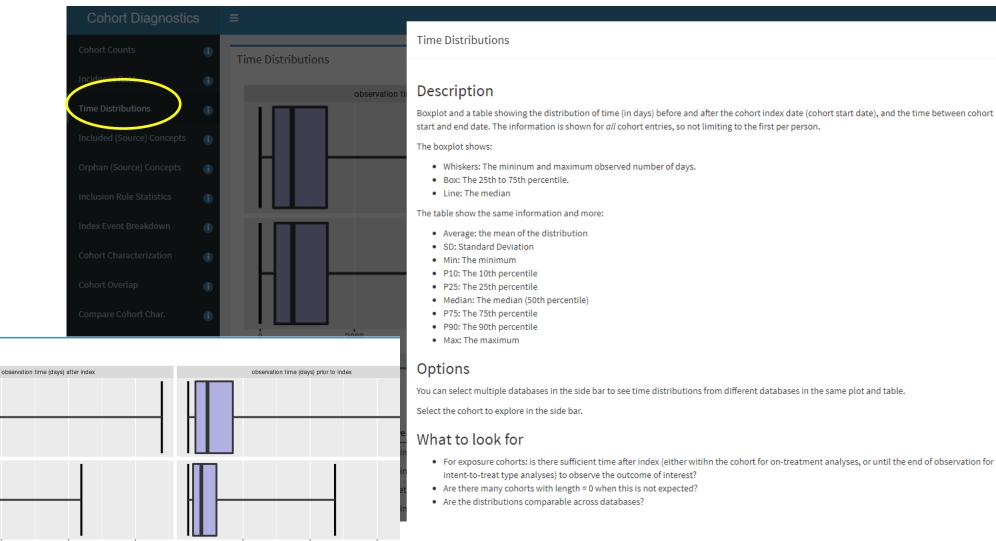
Use same y-scale across databases





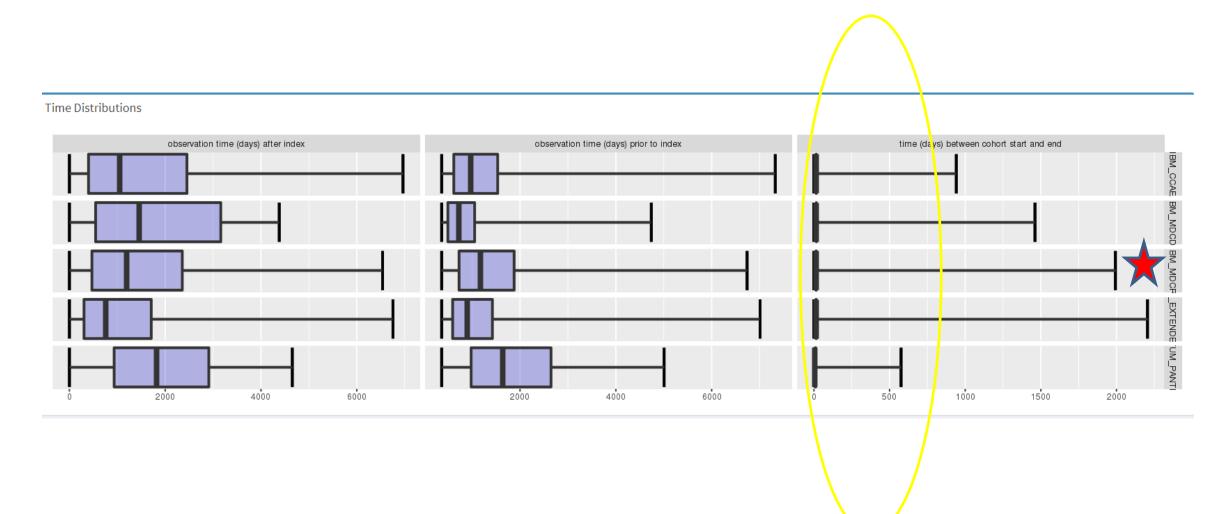
Time Distributions

Functionality – Time distributions





Functionality – Time distributions





Functionality – Included concepts

	Cohort Diagnostics	=							
				Included (Source) Concepts					
		Source Concepts O Standard Concept		pt					
		Show 25 🗸 entries		Description					
	Time Distributions	Subjects 🔷 Co	oncept ID 🔶 Voca						
	Included (Source) Concepts 👔	14,145,545	44821949 ICD9	concepts are identified in the _source_concept_id fields of the Common Data Model, (e.g. drug_source_concept_id) and are used to identify the					
	Orphan (Source) Concepts	11,014,812 7,035,441	44834715 ICD9 35207668 ICD1	specific source codes used in a database. Standard concepts are found using the _concept_id fields (e.g. drug_concept_id), and use the same coding system across all databases.					
		1,021,428	44823109 ICD9						
		191,229 44833715 ICD9		Options					
		153,724	44836077 ICD9	Select the cohort and the specific concent set within that cohort to explore in the side har					
		132,233	44824390 ICD9 44822082 ICD9	75 m					
		104,919 99,895	44822082 1CD9						
		97,646	44822084 ICD9						
		76,342	44830637 ICD9	account:					
		74,071	45563052 ICD1	codes present?					
	Database	65,693	44823229 ICD9 44825589 ICD9C						
Subjects 🔶	Concept ID 🔶 Vocabula		Name	y hypertension, unspecified					
	• •		-	sive urgency					
14,145,545	44821949 ICD9CM	401.9	Unspecifie	fied essential hypertension al [pregnancy-induced] hypertension without significant proteinuria, unspecified trimester					
11,014,812	44834715 ICD9CM	401.1	Benign ess	essential hypertension ed maternal hypertension, third trimester					
7,035,441	35207668 ICD10CM	110	Essential (l (primary) hypertension					
1,021,428	44823109 ICD9CM	401.0	Malignant	nt essential hypertension					



2,211,273

978,841

35211268

44837038

ICD10CM

ICD9CM

R03.0

365.04

Functionality – Orphan concepts

	Cohort Diagnostics									
Co		()	Show 25 💙 entries		Orphan (Source) Concepts					
Inc		i	Count 🔶	Concept ID 🔶 Standa						
			6,669,185	141693 S	Description					
Tir		()	4,457,912	44829313	A table showing the (source) concepts observed in the database that are not included in a concept set of a cohort, but maybe should be. The					
Inc			2,9 <mark>38,062</mark>	381290 S	following logic is used to indentify concepts that might be relevant:					
		<u> </u>	2,211 <mark>,273</mark>	35211268	1. Given a concept set expression, find all included concepts.					
Qr	phan (Source) Concevits	i	978,8 <mark>41</mark>	44837038	 Find all names of those concepts, including synonyms, and the names of source concepts that map to them. Search for concepts (standard and source) that contain any of those names as substring. 					
Les.	ducion Dulo Statistica		810,39 <mark>8</mark>	44782429 S	4. Filter those concepts to those that are not in the original set of concepts (i.e. orphans).					
Inc		•	620,481	44782690 S	5. Restrict the set of orphan concepts to those that appear in the CDM database as either source concept or standard concept.					
Inc		(i)	476,461	4013643 S	The Subjects column contains the number of subjects in the entire database that have the specific concept. This count is not restricted to only those people in the cohort. Source concepts are identified in the _source_concept_id fields of the Common Data Model, (e.g. drug_source_concept_id) and					
			378,036	312902 S	identify the specific source codes used in a database. Standard concepts are found using the _concept_id fields (e.g. drug_concept_id),					
Co		()	377,747	44823113	and use the same coding system across all databases.					
Co		•	363,781	4064925 S	Options					
		Ŭ	363,781	44832016						
Co		()	254,253	4313767 S	You can select a database in the side bar to see the concepts and counts observed in that database.					
De			251,673	45543044	Select the cohort and the specific concept set within that cohort to explore in the side bar.					
Da			242,939	44833447	What to look for					
Dat	tabase		234,058	192680 S	Are there concepts that are not included in the concept but should be? Note that the provided list likely contains many false positives.					
	0015		100.150	11700715 0	• Are there concepts that are not included in the concept but should be: Note that the provided list likely contains many latse positives.					
nt 🔷 ,185	Concept ID Standard 141693 S	Vocab SNOM		Name	blood-pressure reading without diagnosis of hypertension Secondary pulmonary hypertension					
,185	44829313	ICD9C			blood pressure reading without diagnosis of hypertension blood pressure reading without diagnosis of hypertension					
,062	381290 S	SNOM			/pertension					

Elevated blood-pressure reading, without diagnosis of hypertension

Ocular hypertension

20

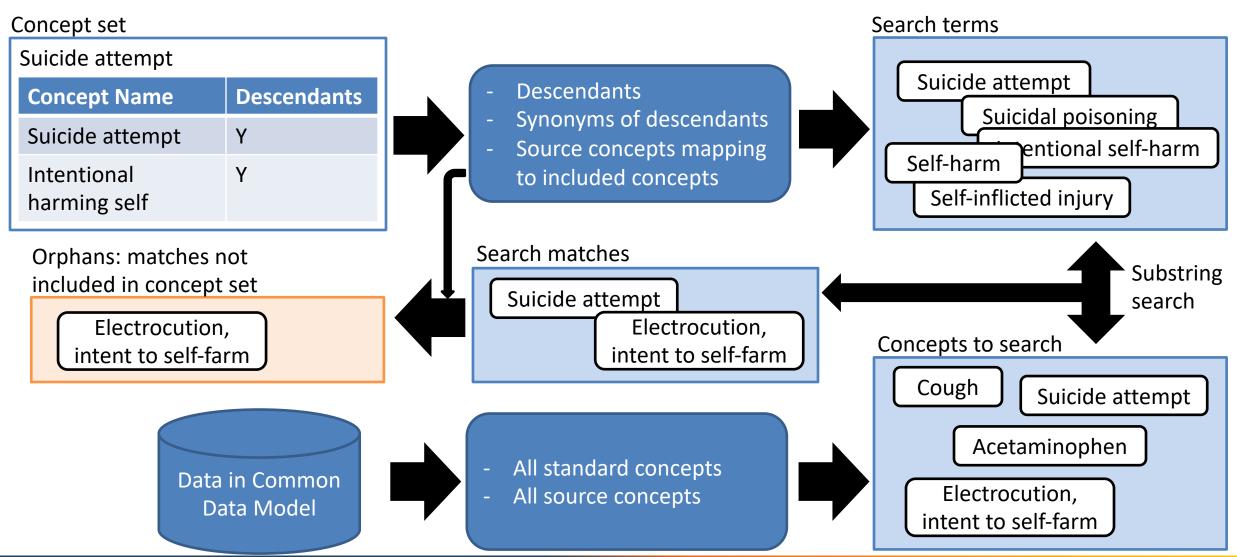


Functionality – Orphan concepts

Cohort Diagnostics	≡				
Cohort Counts	Show 25 🗸 entries				
Incidence Rate	Count ≑	Concept ID 🍦 Standard	🔶 Vocabulary	🔶 Code	Vame
	9,982,695	80180 S	SNOMED	396275006	Osteoarthritis
Time Distributions	5,172,937	44831489	ICD9CM	715.90	Osteoarthrosis, unspecified whether generalized or localized, site unspecified
Included (Source) Concepts	3,670,692	4079750 S	SNOMED	239873007	Osteoarthritis of knee
	3,554,961	75617 S	SNOMED	201819000	Degenerative joint disease involving multiple joints
Orphan (Source) Concepts	3,298,489	45586894	ICD10CM	M19.90	Unspecified osteoarthritis, unspecified site
Inclusion Rule Statistics	2,676,722	4035439 S	SNOMED	239862000	Idiopathic osteoarthritis
	2,517,379	44836171	ICD9CM	715.00	Osteoarthrosis, generalized, site unspecified
Index Event Breakdown	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
Cohort Characterization	1,104,621	35208765	ICD10CM	M15.9	Polyosteoarthritis, unspecified
Cohort Overlap	1,037,582	44829195	ICD9CM	715.09	Osteoarthrosis, generalized, multiple sites
	954,396	72993 S	SNOMED	201829007	Localized, primary osteoarthritis
Compare Cohort Char.	810,869	35208772	ICD10CM	M17.0	Bilateral primary osteoarthritis of knee
Database information	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
Database	621,026	72990 S	SNOMED	90860001	Localized osteoarthrosis uncertain if primary OR secondary
IBM_MDCD 🗸	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
Cohort (Target)	574,389	45538681	ICD10CM	M17.12	Unilateral primary osteoarthritis, left knee
[Cohort Diagnostics] Rheu 🛛 🗸	573,064	73840 S	SNOMED	201831003	Localized, primary osteoarthritis of the shoulder region
	546,943	35208776	ICD10CM	M17.9	Osteoarthritis of knee, unspecified
Concept Set	520,222	44829197	ICD9CM	715.89	Osteoarthrosis involving, or with mention of more than one site, but not specified as generalized, multiple sites
Rheumatoid arthritis 🛛 🗸	384,098	4035611 S	SNOMED	239791005	Seropositive rheumatoid arthritis
	379,690	44834980	ICD9CM	715.95	Ostcoarthrosis, unspecified whether generalized or localized, pelvic region and thigh
	379,690	40483794 S	SNOMED	445478004	Degenerative joint disease of pelvis



Orphan codes





Functionality – Inclusion Rule Statistics

Cohort Counts	i	Show 25 🗸 entries
Incidence Rate		Sequence 🔶 Nar
Time Distributions		0 has 1 Has
Included (Source) Concepts		2 Is o
Orphan (Source) Concepts		Showing 1 to 3 of 3 entries
Inclusion Rule Statistic		
Index Event Breakdown		
Cohort Characterization		
Cohort Overlap		
Compare Cohort Char.		
Database information		
Database		
CCAE	~	
Cohort (Target)		
New users of ACE inhibitor	~	
venuence Name		

Inclusion Rule Statistics

Description

A table showing the number of subject that match specific inclusion rules in the cohort definition. Note that this table will be empty if no inclusion rules have been specified.

The table contains the following columns:

- · Sequence: The order in which the inclusion rules are applied to the cohort.
- Name: The name of the inclusion rule.
- Meet: The number of cohort entries that meet the entry event definition and the specific inclusion rule indicated in the row.
- Gain: The number of cohort entries that would be gained if this inclusion rule was dropped.
- Total: The number of cohort entries meeting the entry event definition. In other words, the number of cohort entries before applying any of the inclusion rules.
- Remain: The number of cohort entries remaining after applying the specific inclusion rule, and all preceding rules.

Options

You can select a database in the side bar to see the inclusion rule statistics observed in that database.

Select the cohort to explore in the side bar.

What to look for

- · Are there inlusion rules that nobody meets in a database? For example, requiring a specialist visit that isn't recorded in a specific database.
- Are there inclusion rules that have no effect in a database? For example, requiring no occurrence of a prior disease code that is not recorded in a database.
- Are there inclusion rules that drastically reduce the population? In this case we might worry about generalizability. For example, if we require a
 diagnostic procedure, and only a small fraction meets this criteria, we may wonder if this identifies a special population that differs from the
 overall population in significant ways.

Sequence	Name	÷ N	leet 🔶	Gain 🌲	Total 🔶	Remain 🔶
(0 has hypertension diagnosis in 1 yr prior to treatment	2,54	47,647	539,522	3,604,155	2,547,647
1	1 Has no prior antihypertensive drug exposures in medical history	1,90	67,168	929,300	3,604,155	1,241,966
2	2 Is only taking ACE as monotherapy, with no concomitant combination treatments	2,5	55,382	395,771	3,604,155	846,195



Functionality – Index Event Breakdown

Conort Diagnostic	=					
	Show 25 v entries					
	Concept ID					
	19080128					
	19080129					
	1308251					
	1308250					
	1308221					
	40171661					
inclusion rule statistics	40165789					
Index Event Breakdown	40165762					
	40171671					
	19003855					
	1334460					
	19003854					
	19067557					
	1334459					

Index Event Breakdown

Description

A table showing the concepts belonging to the concept sets in the entry event definition that are observed on the index date. In other words, the table lists the concepts that likely triggered the cohort entry. The counts indicate number of cohort entries where the concepts was observed on the index date. Note that multiple concepts can be present on the index date, so the sum of counts might be greater than the cohort entry count.

Options

You can select multiple databases in the side bar to see counts from different databases side-by-side.

Select the cohort to explore in the side bar.

What to look for

- Is one concept unexpectedly dominating? For example, if our cohort identies exposure to drugs in a class, but we notice almost everyone enters the cohort based on a single drug, we may wonder whether our results will generalize to the class.
- Are the highest ranking concepts different across databases? For example, is everyone in one database initiating high-dose prescriptions, and everyone in another database low-dose prescriptions?

1334459 Ramipril 2.5 MG Oral Capsule

	40165772 Englanril Maleate 20 MG Oral Tablet		
Concept ID 🍦 Name	÷	CCAE Count 🔷	MDCD Count 🔷
19080128 Lisinopril 10 MG Oral	ablet	385,468	37,426
19080129 Lisinopril 20 MG Oral	ablet	154,099	16,531
1308251 Lisinopril 5 MG Oral T	blet	127,815	16,356
1308250 Lisinopril 2.5 MG Oral	Tablet	23,4 <mark>8</mark> 7	4,189
1308221 Lisinopril 40 MG Oral	ablet	16,166	2,678



Functionality – Index Event Breakdown

Show 25 V entries												
Concept ID 🔶 Name	CPRD Count	IBM_CCAE Count 🔶	IBM_MDCD Count 🔶	IBM_MDCR Count 🔶	JMDC Count 🔶	OPTUM_EXTENDED_DOD Count 崇	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	28,77 <mark>5</mark>	72,288					
4083556 Seronegative rheumatoid arthritis	6 , 86 <mark>6</mark>	18,587	3,700	3,011	374	28,775	72,288					
4035611 Seropositive rheumatoid arthritis	2,280	26,030	6,443	3,736	108	45,1 <mark>5</mark> 9	127,277					
4035611 Seropositive rheumatoid arthritis	2,280	26,030	6,443	3,736	108	45,1 <mark>5</mark> 9	127,277					
4114444 Flare of rheumatoid arthritis	710						111					



Pretty O Raw
 Covariate Name
 Age group
 10-14
 15-19
 20-24
 25-29
 30-34
 35-39
 40-44
 45-49
 50-54
 55-59
 60-84
 65-69

Gender: female Medical history: General

Crohn's disease Dementia

Acute respiratory disease

Attention deficit hyperactivity disorder Chronic liver disease

Chronic obstructive lung disease

Functionality – Cohort Characterization

Cohort Diagnostic	s	=	
Cohort Counts			Cohort Characterization
Incidence Rate		Pretty	
		Covariate Name	Description
Time Distributions		Age group	A table showing cohort characteristics (covariates). These characteristics are captured on or before the cohort start date. There is a Pretty and a Raw version of this table.
Included (Source) Concepts		10-14	The Pretty table shows a few selected covariates. These are all binary covariates, and the table shows the proportion (%) of the cohort entries having the covariate.
Orphan (Source) Concepts		15-19	
Inclusion Rule Statistics		20-24	The Raw table shows all captured covariates. These include binary and continuous covariates (e.g. the Charlson comorbidity index). For each covariate the table lists the mean, which for binary covariates is equal to the proportion, and the standard deviation (SD).
Index Eve <u>nt Brea</u> kdown		25-29 30-34	Options
Cohort Characterization		35-39	You can select multiple databases in the side bar to see cohort characteristics from different databases side-by-side in the same table.
Conort Characterization		40-44	Select the cohort to explore in the side bar.
Cohort Overtap		45-49	Select either the Pretty or the Raw table at the top of the table.
Compare Cohort Char.		50-54	What to look for
		55-59 ссле р	 Are the characteristics of the cohort as expected? For example, do people have the expected comorbidities? Do the characteristics of the cohort differ much per database?
			0.2%
			0.7% 1.4%
			5.0
			8,2% 12,0%
			16.0% 18.6%
			18.4%

1.3% 38.3%

25.2%

1.2%

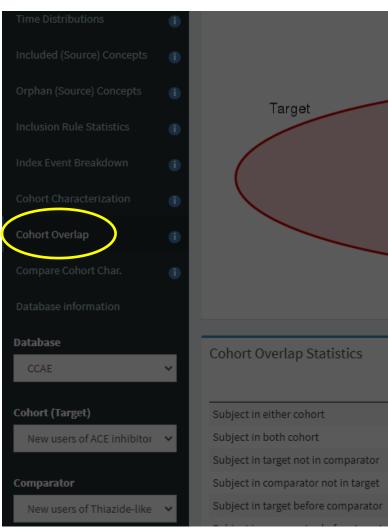
1.4% 1.6%

0.3%

0.1%

26

Functionality - Cohort Overlap (Subjects)

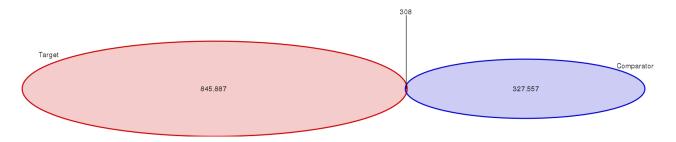


A Venn diagram showing the overlap between two cohorts, and a table listing several overlap statistics.

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:

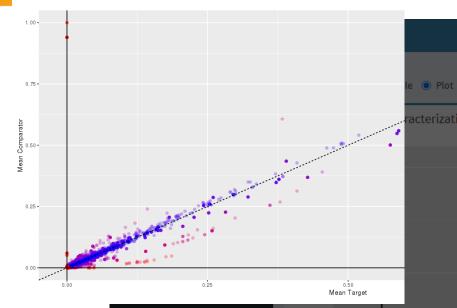
- Subject in either cohort: The number of subjects that enter one or both cohorts. (The union)
- Subject in both cohort: The number of subjects that enter both cohorts, although not necessarily at the same time. (The intersection)
- Subject in target not in comparator: The number of subjects that enter 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



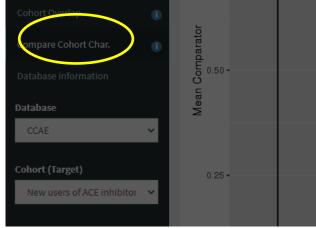
Functionality - Cohort Overlap (Subjects)



Functionality – Compare Cohort Characteristics



Compare Cohort Characterization



Compare Cohort Characteristics

Description

A table or plot showing cohort characteristics (covariates) for two cohorts side-by-side. These characteristics are captured on or before the cohort start date. There is a Pretty and a Raw version of the table.

The Pretty table shows a few selected covariates. These are all binary covariates, and the table shows the proportion (%) of the cohort entries having the covariate, as well as the standardized difference of the mean (StdDiff).

The Raw table shows all captured covariates. These include binary and continuous covariates (e.g. the Charlson comorbidity index). For each covariate the table lists the mean, which for binary covariates is equal to the proportion, the standard deviation (SD), and the standardized difference of the mean (StdDiff).

The plot shows all covariates, include binary and continuous covariates. The x-axis represents the mean value in the target cohort, the y-axis the mean value in the comparator cohort. Each dot represents a covariate, and the color indicates the absolute value of the standardized difference of the mean.

Options

You can select a database in the side bar.

Select the cohort to explore in the side bar.

Select either the Pretty, the Raw table, or the plot at the top of the screen.

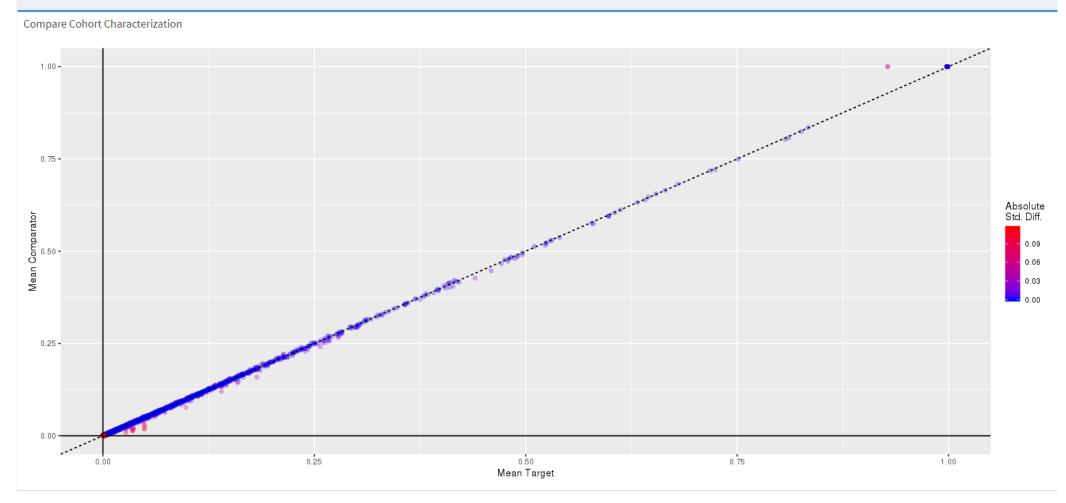
In the plot, you can move the mouse pointer over a dot to see information on that covariate.

What to look for

- Are there major differences between the two cohorts? For example, if we wish to compute a propensity score between two cohorts, concepts that have very high proportion in one cohort and a very low proportion in the other may lead to a perfectly predictive model.
- In general, how comparable are two cohorts? If we wish to compare two exposures, but the cohorts differ over many characterics, we may be able to fit a propensity model and compute an estimate, but we may have concerns over the generalizability of the results.

Functionality – Compare Cohort Characteristics

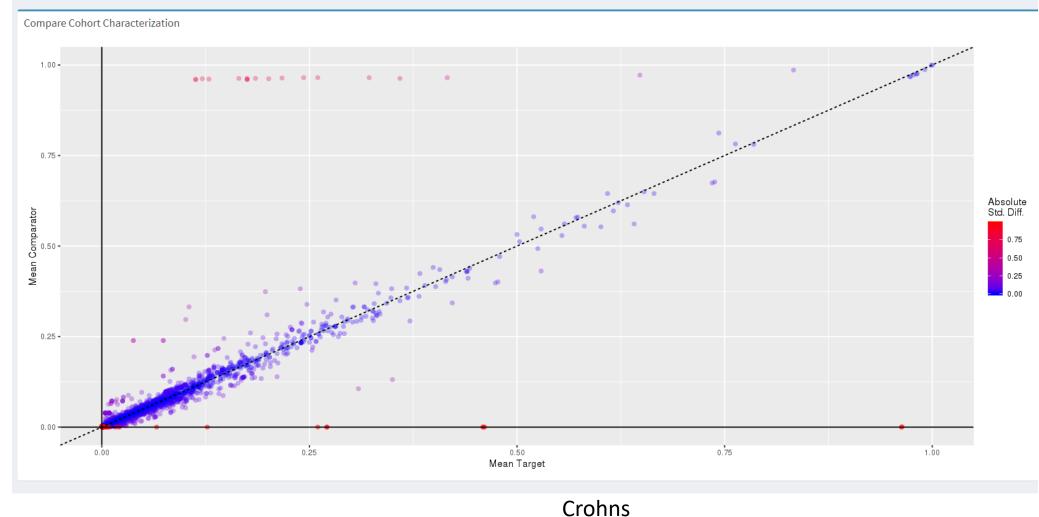
○ Pretty table ○ Raw table ● Plot





Functionality – Compare Cohort Characteristics

○ Pretty table ○ Raw table ● Plot



Ulcerative colitis

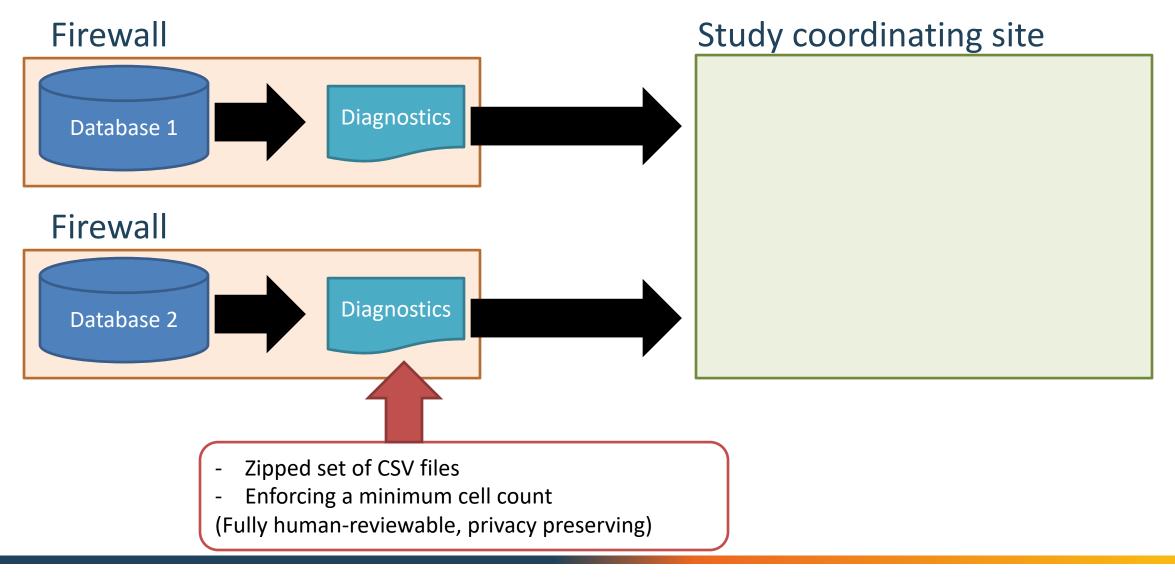


Benefits of Cohort Diagnostics

- 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 \rightarrow software functionality



Walk thru Use case

Backup



Clinical review

© 2020 UpToDate, Inc. and/or its affiliates. All Rights Reserved.

Gell and Coombs classification of immunologic drug reactions

Туре		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 vasoactive substances, such as histamine, prostaglandins, and leukotrienes.		Angioede	aphylaxis gioedema onchospasm	
II	- S - J	plex syndrome – most kin manifestations oint manifestations ever	ly clinical diagnosis (no testing)		(hives) ion c anemia cytopenia nia 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).	Some mo Severe e AGEP DRESS/D Interstiti Drug-ind	dermatitis orbilliform reactions xfoliative dermatoses (eg, SJS/TEN) DiHS al nephritis uced hepatitis esentations	

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/druginduced hypersensitivity syndrome.

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



Data review

Colum	n visibility Copy	CSV Show 15	entries					Filter: Search
Showin	g 1 to 4 of 4 entries							Previous 1 Next
1	ld 🍦	Code 🔶	Name	Class	♦ RC♦	DRC	Domain	🔷 Vocabulary 🔶
1	4297831	403608009	Serum sickness due to drug	Clinical Finding	188	188	Observation	SNOMED
1	4297825	402658008	Serum sickness type vasculitis	Clinical Finding	36	36	Condition	SNOMED
1	40410225	213322006	Serum sickness	Clinical Finding	0	0	Condition	SNOMED
1	40410227	213324007	Serum sickness	Clinical Finding	0	0	Condition	SNOMED

Showing 1 to 4 of 4 entries

Not coded in data!!

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

Previous 1 Next

Prevalence	
Treemap	Table
Column visib	ility 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
L	Column visibility Copy CSV Show 15 V 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)

Hierarchy

♠ Parents								
	Column visibility Copy CSV Show 15 T entries						Filten	Search
	Showing 1 to 2 of 2 entries							Previous 1 Nex
	Id Code Name		Class	RC	DRC	Distance	Domain	Vocabulary
Yocabulary SNOMED (2)	F 4231425 89322006 Urticaria medicamentosa		Clinical Finding	262	450	1	Observation	SNOMED
▼ Class Clinical Finding (2)	🔚 4081075 238697009 Immune complex urticaria		Clinical Finding	0	224	1	Condition	SNOMED
Has Records true (1) false (1)	Showing 1 to 2 of 2 entries							Previous 1 Ne
▼ Has Descendant Records true (2)								
➤ Current Concept								
Showing 1 to 1 of 1 entries								
ld 🕴 Code 🍦 Nan	ne	Class	RC	DRC	Distar	ice 🔻 🛙	omain 🕴	Vocabulary
4297831 403608009 Sen	Clinical Finding	188	188	0	0	Observation	SNOMED	



Literature review

AMERICAN JOURNAL OF EPIDEMIOLOGY Copyright © 1990 by The Johns Hopkins University School of Hygiene and Public Health All rights reserved Vol. 132, No. 2 Printed in U.S.A.

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 PLATT¹

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%	<mark>0.14%</mark>		
Penicillin V	0.0059%	0.04%		

*20 days of starting antibiotic Observational data

Case definition:

In 1990!

1) Urticaria, hives, rash (including erythema) multiforme), or angioedema

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		<mark>0.02%</mark>
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

how 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

pehr



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

Itch, Urticaria, Acute Rash	
Cohort Entry Events	
Events having any of the following criteria	a:
a condition occurrence of [EPI_767-O]	Skin manifestations 👻
an observation of [EPI_767-O] Skin ma	anifestations 👻
	▼ days before and 0 ▼ days after event index date er person.
Inclusion Criteria	
New inclusion criteria	[EPI_767-O] Joint manifestations (broad) -any musculoskeletal pain not including arthropathy, muscle pain
 [EPI_767-O] Joint manifestations (broad) -any musculoskeletal pain not including arthropathy, muscle pain 	enter an inclusion rule description having all v of the following criteria:
	with at least 1 using all occurrences of: a condition occurrence of [EPI_767-O] Joint manifestation where event starts between 0 days Before and 0 days After index start date add additice The index date refers to the event from the Cohort Entry criteria.



Composite definition

🔲 Age 🔲 Gender 🗹 Calendar Year

Use same y-scale across databases

