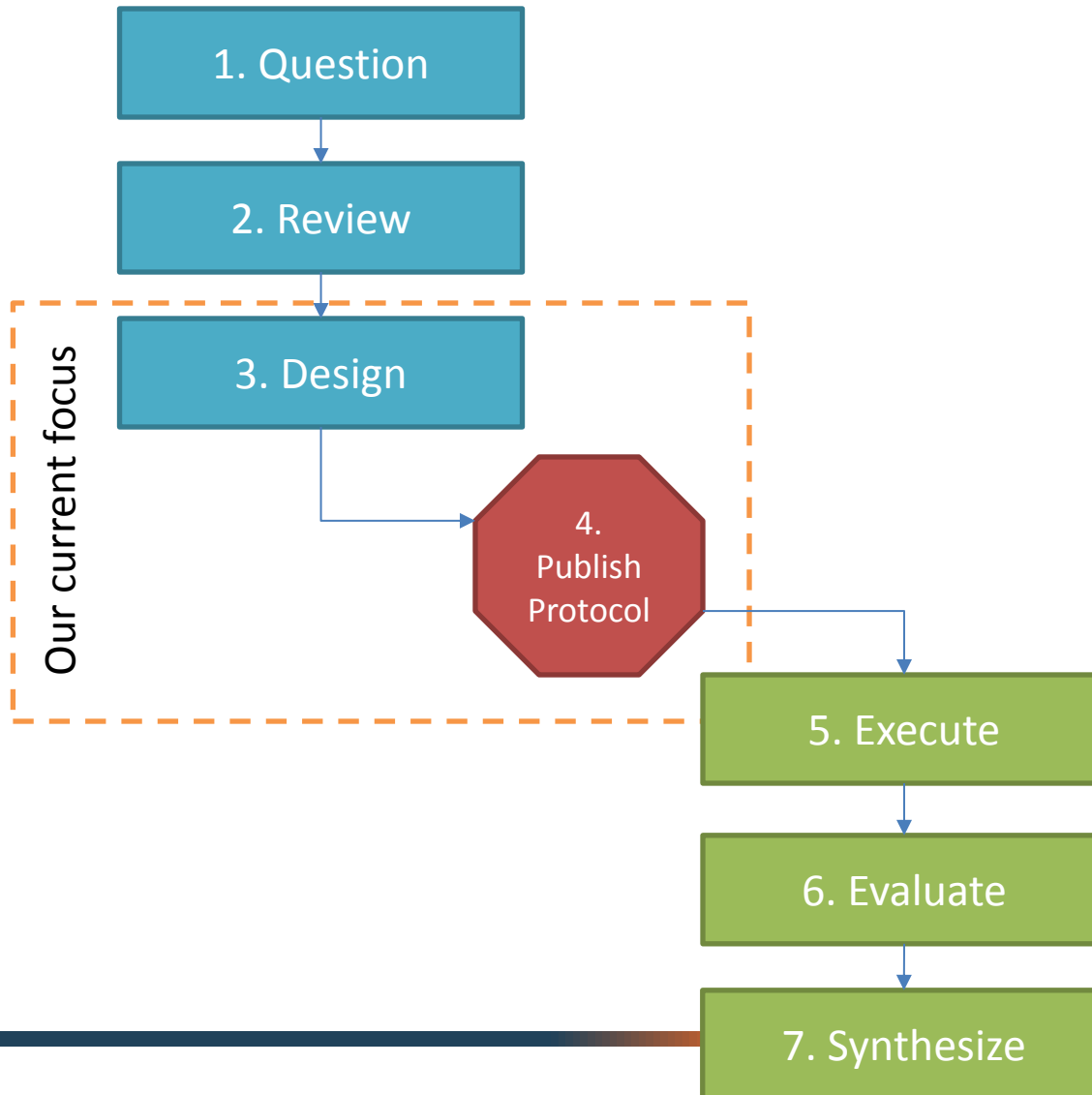




Study designer track:  
Deep dive into cohort study  
design using ATLAS



# A standardized process for evidence generation and dissemination





# What is a protocol?

“In the natural sciences a protocol is a **predefined written procedural method** in the design and implementation of experiments.

Protocols are written whenever it is desirable to standardize a laboratory method to **ensure successful replication of results** by others in the same laboratory or by other laboratories.

Detailed protocols also facilitate the assessment of results through peer review.”



# What should a protocol be in epidemiology?

“The study protocol is a core document of a study. A protocol should be drafted as one of the first steps in any research project. The final version **must precisely describe everything being done in the study so that the study can be reproduced.**”

-ENCEPP Guide on Methodological Standards in Pharmacoepidemiology

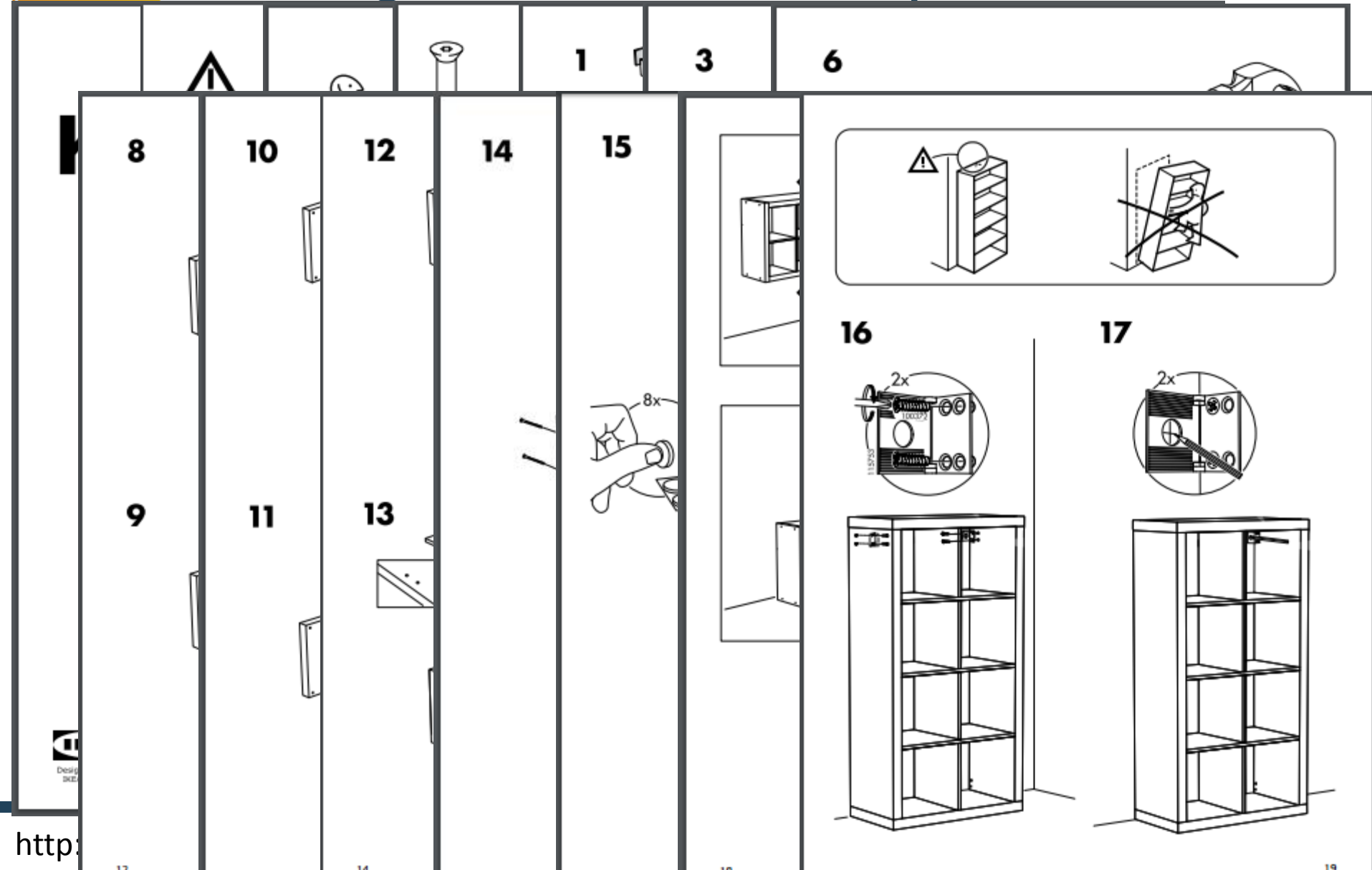


# Analogy for a retrospective analysis of observational healthcare data

- Assembly Instructions:
  - Initial parts already exist
  - Defined step-by-step procedure can be followed
  - If followed correctly, you should always get the same output



# Lessons from IKEA: making effective assembly instructions





# Protocol / assembly instructions: initial warnings



## ENGLISH

Important information  
Read carefully.  
Keep this information for further reference.

## WARNING

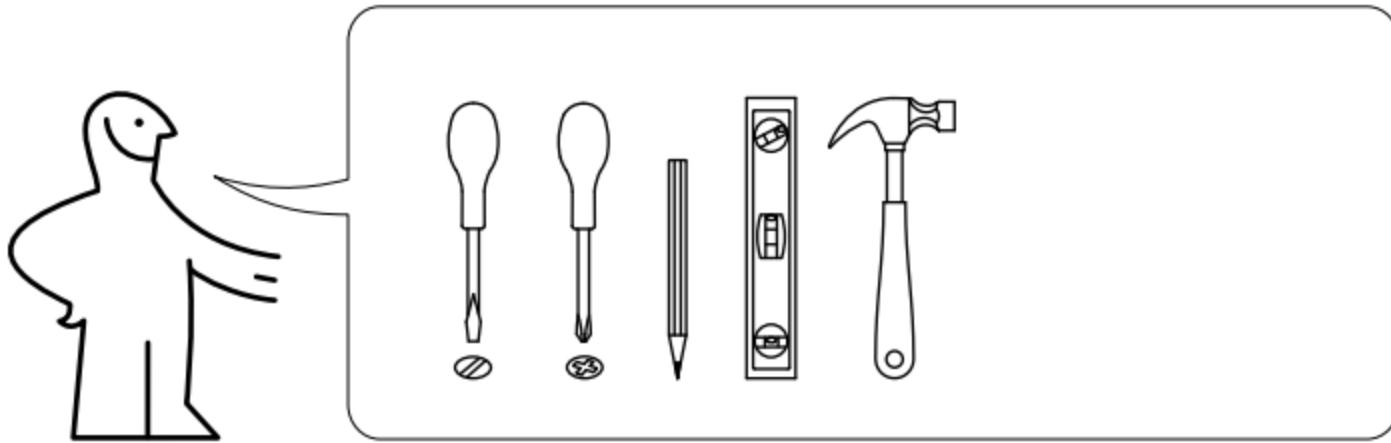
**Serious or fatal crushing injuries can occur from furniture tip-over. To prevent this furniture from tipping over it must be permanently fixed to the wall.**

Fixing devices for the wall are not included since different wall materials require different types of fixing devices. Use fixing devices suitable for the walls in your home. For advice on suitable fixing systems, contact your local specialized dealer.

- What warnings should be communicated in your epidemiology protocol, prior to execution of the analysis?



# Protocol / assembly instructions: required tools

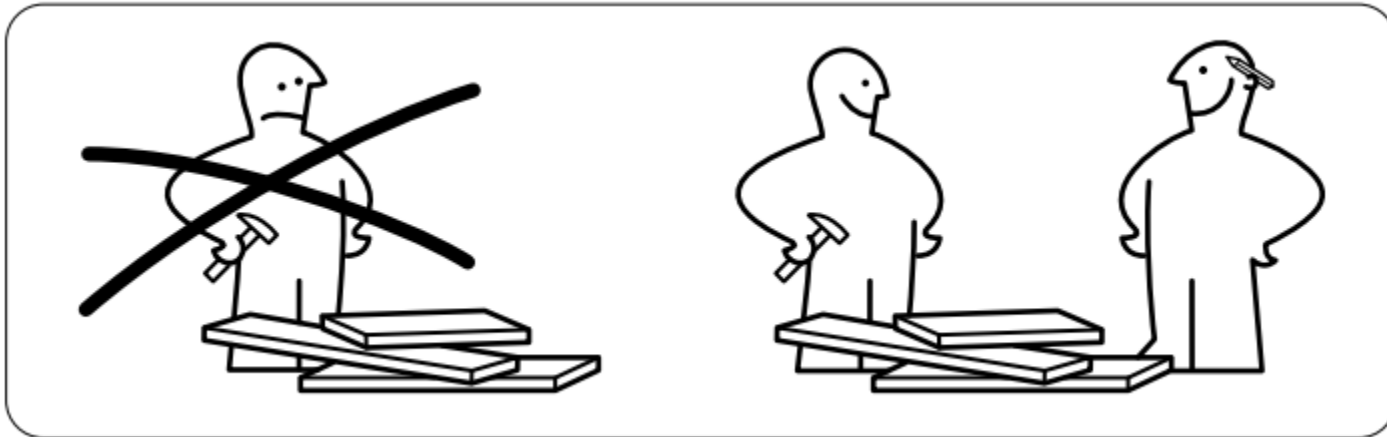


- What tools do you need to have access to in order to properly complete the analysis?





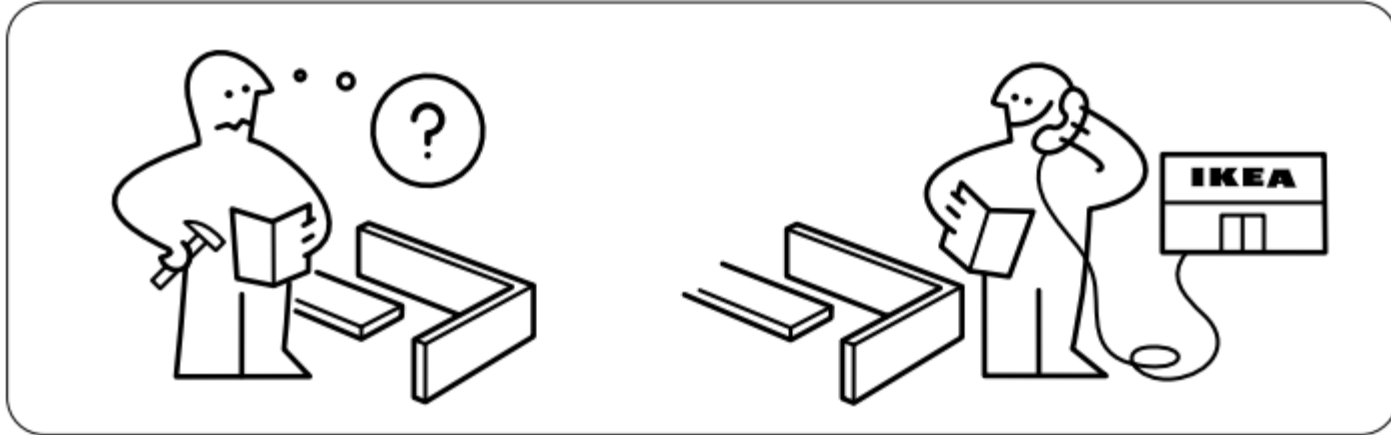
# Protocol / assembly instructions: required expertise



- Study design should not be a one-person effort
- Anyone in any role can contribute an initial research question...
- ...but it's unlikely that anyone has all of the necessary expertise to design and implement a study to answer that question
  - Therapeutic area and clinical domain knowledge
  - Understanding how the clinical phenomena manifests in the health care system and data capture processes
  - Working competency with the observational databases and source vocabularies
  - Expertise with standardized analytics tools to design and implement analyses
- Protocol development should be shared collaboration activity



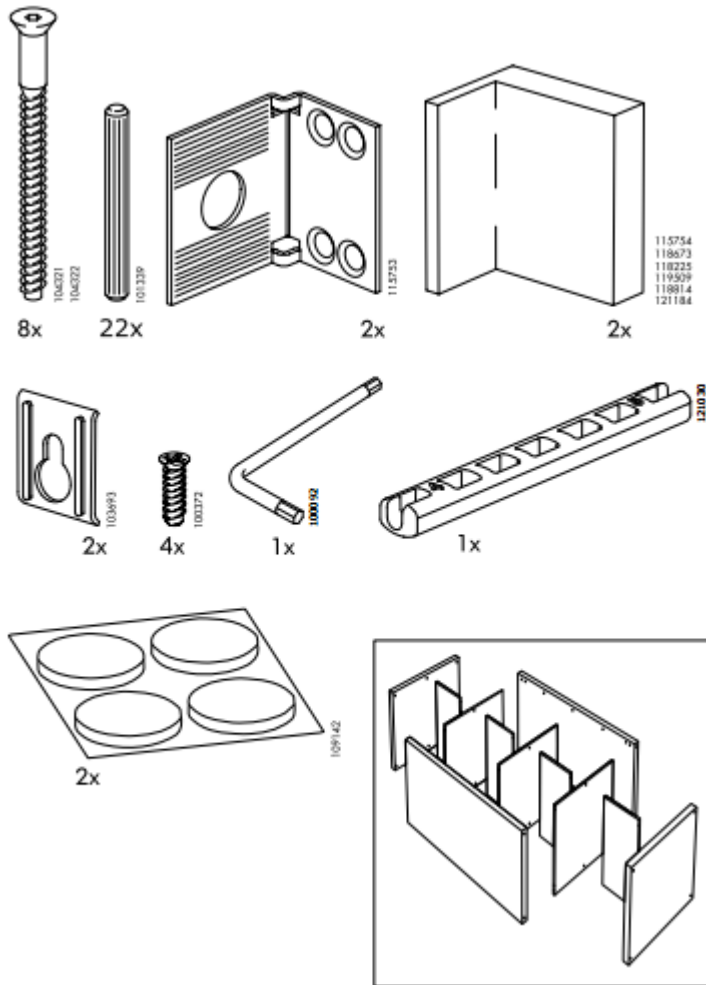
# Protocol / assembly instructions: required expertise



- A protocol should be a complete specification of all procedures to be executed, which is pre-defined and documented prior to study execution
- There should be a pre-defined process for how to reconcile any ambiguities identified during implementation
- All resolutions should be documented as a protocol amendment



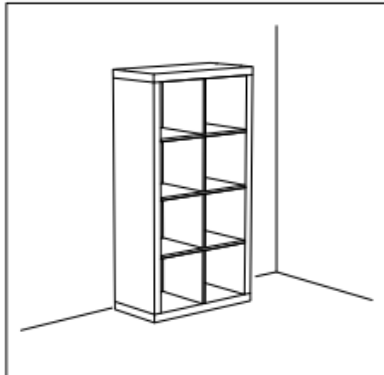
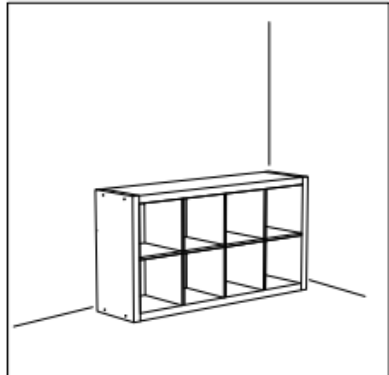
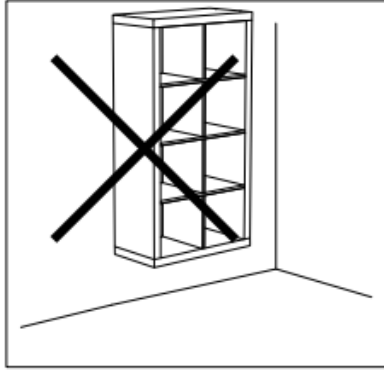
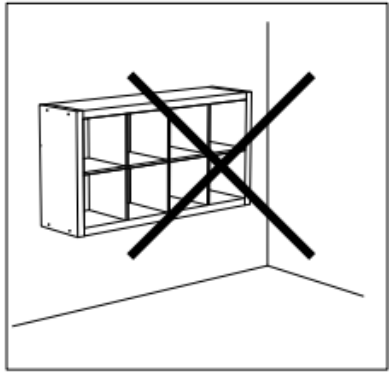
# Protocol / assembly instructions: complete inventory of initial inputs



- IKEA doesn't just say: 'use some particle board and a few screws'....so we shouldn't just say: 'use an administrative claims database'
- Full specification requires documenting scope of source data used (release date, scope of calendar time and population), version of CDM and ETL process, version of vocabulary



# Protocol / assembly instructions: intended use for the final outputs



- It is important to decide upfront how the evidence generated it going to be used once completed
  - What decision is being made?
  - Who is the decision-maker?
  - How can the evidence this study generates inform the decision?



# Observational data analysis is a science, not an art

- If you want to generate reliable evidence from observational healthcare databases to meaningfully inform medical decision making, you must apply a consistent, reproducible, verifiable process that follows the scientific method
- If you are looking for ‘artistic’ opportunities:
  - Propose interesting and clinically relevant questions to answer
  - Discover new data capture processes and incorporate these elements into the OMOP common data model
  - Design novel visualizations to more effectively communicate the evidence generated



# Replication of Garbe et al. using the OHDSI framework

Eur J Clin Pharmacol (2013) 69:549–557

DOI 10.1007/s00228-012-1334-2

---

PHARMACOEPIDEMIOLOGY AND PRESCRIPTION

## **High-dimensional versus conventional propensity scores in a comparative effectiveness study of coxibs and reduced upper gastrointestinal complications**

E. Garbe • S. Kloss • M. Suling • I. Pigeot •  
S. Schneeweiss

Received: 28 February 2012 / Accepted: 6 June 2012 / Published online: 5 July 2012

© Springer-Verlag 2012

---



# What is the design used by Garbe et al?

Input parameter	Design choice
Target cohort (T)	Celecoxib new users
Comparator cohort (C)	Traditional non-steroid antiinflammatory drugs (NSAID) new users
Outcome cohort (O)	Upper gastrointestinal complications (UGIC)
Time-at-risk	cohort start → cohort end
Model specification	1:1 propensity score-matched multivariable conditional Poisson regression



# Garbe et al. description of cohort(s)

This was a cohort study in which a new user design was applied. The aim was to estimate the effect of tNSAIDs and coxibs on the risk of UGIC. New users of tNSAIDs or coxibs were defined as patients who were continuously enrolled in their SHI provider for at least 12 months without any notation of NSAID use, including coxibs, during this time period. Cohort entry was the first notation of a prescription for a tNSAID or a coxib. Cohort exit was defined as discontinuation or switch of the initial NSAID, disenrollment from the SHI provider, hospitalization for UGIC, hospital diagnosis of cancer, death, or the end of the study period, whichever came first. Patients were required to be at least 16 years of age at the time of first use and not to have a diagnosis of cancer in the 12 months preceding cohort entry.





# Garbe et al. replication: Implementing the target cohort in ATLAS

ATLAS x hix.jnj.com/atlas/#/cohortdefinition/2727 ryan@ohdsi.org

ATLAS Cohort

OHDSI estimation tutorial: Garbe replication: target cohort - celecoxib new users Save Close Copy

Definition Concept Sets Generation Reporting Explore Export

**Cohort definition:** A cohort is defined as the set of persons satisfying one or more inclusion criteria for a duration of time. One person may qualify for one cohort multiple times during non-overlapping time intervals. Cohorts are constructed in ATLAS by specifying cohort entry criteria and cohort exit criteria. Cohort entry criteria involve selecting one or more initial events, which determine the start date for cohort entry, and optionally specifying additional inclusion criteria which filter to the qualifying events. Cohort exit criteria are applied to each cohort entry record to determine the end date when the person's episode no longer qualifies for the cohort.

All Cohort Entry Criteria Cohort Exit Criteria

**Initial event cohort:** Events are recorded time-stamped observations for the persons, such as drug exposures, conditions, procedures, measurements and visits. All events have a start date and end date, though some events may have a start date and end date with the same value (such as procedures or measurements). The event index date is set to be equal to the event start date.

People having any of the following: Add Initial Event...

a drug era of celecoxib Add

✗ for the first time in the person's history

✗ era start is: Between 2004-01-01 and 2007-12-31

✗ with age at era start Greater or Equal To 16

with continuous observation of at least 365 days before and 0 days after event index date

Limit initial events to: earliest event per person.

**Index event inclusion criteria:** From among the index events, include:

People having all of the following criteria: Add New Criteria...

with exactly 0 using all occurrences of:

a condition occurrence of Broad malignancies excluding skin cancer Add

occurring between 365 days Before and 0 days Before event index date

and with exactly 0 using all occurrences of:

a drug exposure of NSAIDs Add

occurring between 365 days Before and 1 days Before event index date

Delete Criteria

Delete Criteria



# Implementing the target cohort in ATLAS: Defining the initial event

**Initial event cohort:** Events are recorded time-stamped observations for the persons, such as drug exposures, conditions, procedures, measurements and visits. All events have a start date and end date, though some events may have a start date and end date with the same value (such as procedures or measurements). The event index date is set to be equal to the event start date.

People having any of the following: [Add Initial Event...](#)

a drug era of  [Add](#) [Add criteria attribute...](#) [Delete Criteria](#)

✗ for the first time in the person's history

✗ era start is:   and

✗ with age at era start

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

Limit initial events to:  per person.

Ambiguities from publication which require detailed specification in a complete protocol:

- 1) What is the time period for exposure?
- 2) Does exposure need to be first time in history, or only require 12 months prior with no exposure?



# Implementing the target cohort in ATLAS: Specifying initial event inclusion criteria

**Index event inclusion criteria:** From among the index events, include:

People having  of the following criteria:

with   using  occurrences of:

a condition occurrence of

occurring between  days  and  days  event index date

and with   using  occurrences of:

a drug exposure of

occurring between  days  and  days  event index date

Ambiguities from publication which require detailed specification in a complete protocol:

- 1) Does 'at least 12 months without any notation of NSAID use...during this period' mean no exposure any time in prior history or any time in last 12 months?
- 2) How do you define 'diagnosis of cancer'?



# Implementing the target cohort in ATLAS: Select cohort exit criteria

**Cohort Exit Criteria**

Cohort exit criteria based on a fixed time period relative to initial event start or end date:

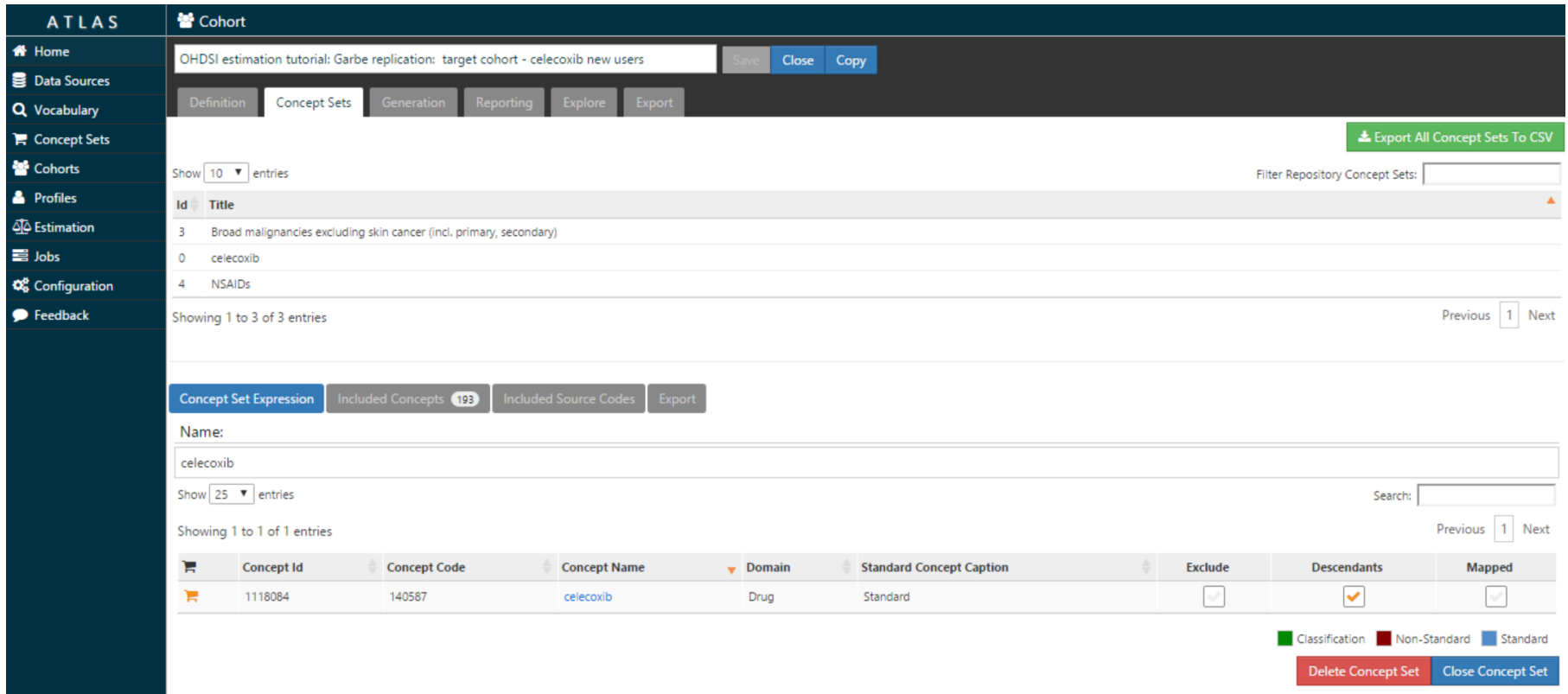
A cohort end date is derived from adding a number of days to be offset from the specified initial event date. If an offset is added to the initial event start date, all cohort episodes will have the same fixed duration (subject to further censoring from other cohort exit criteria). If an offset is added to the initial event end date, persons in the cohort may have varying cohort duration times due to the varying durations of the initial events (such as eras of persistent drug exposure or visit length of stay). This cohort exit criteria assures that the cohort end date will be no greater than the selected index event date, plus the days offset.

- Initial event date to offset from:
- Number of days offset:  days

Ambiguities from publication which require detailed specification in a complete protocol:

- 1) How is continuous exposure defined, such that one can determine a 'discontinuation or switch'?
- 2) How do we differentiate between 'potential time-at-risk' vs. 'realized time-at-risk' to disentangle exposure cohort definition from analytic censoring strategy?

# Implementing the target cohort in ATLAS: Define 'celecoxib' concept set



The screenshot shows the ATLAS Cohort Definition interface. The left sidebar contains navigation links: Home, Data Sources, Vocabulary, Concept Sets, Cohorts, Profiles, Estimation, Jobs, Configuration, and Feedback. The main panel is titled 'Cohort' and shows a breadcrumb trail: 'OHDSI estimation tutorial: Garbe replication: target cohort - celecoxib new users'. Below this, there are tabs for 'Definition', 'Concept Sets', 'Generation', 'Reporting', 'Explore', and 'Export'. The 'Concept Sets' tab is active, displaying a list of concept sets. The first three entries are: 'Broad malignancies excluding skin cancer (incl. primary, secondary)', 'celecoxib', and 'NSAIDs'. The 'celecoxib' concept set is selected, and its details are shown below. The 'Name' field contains 'celecoxib'. The 'Show' dropdown is set to '25' entries. Below this, a table lists the included concepts. The table has columns: Concept Id, Concept Code, Concept Name, Domain, Standard Concept Caption, Exclude, Descendants, and Mapped. The first row shows Concept Id 1118084, Concept Code 140587, Concept Name 'celecoxib', Domain 'Drug', Standard Concept Caption 'Standard', and checked boxes for 'Exclude', 'Descendants', and 'Mapped'. At the bottom right, there are buttons for 'Delete Concept Set' and 'Close Concept Set'.

ATLAS Cohort

OHDSI estimation tutorial: Garbe replication: target cohort - celecoxib new users

Definition Concept Sets Generation Reporting Explore Export

Export All Concept Sets To CSV

Show 10 entries Filter Repository Concept Sets:

Id	Title
3	Broad malignancies excluding skin cancer (incl. primary, secondary)
0	celecoxib
4	NSAIDs

Showing 1 to 3 of 3 entries Previous 1 Next

Concept Set Expression Included Concepts 193 Included Source Codes Export

Name:

celecoxib

Show 25 entries Search:

Showing 1 to 1 of 1 entries Previous 1 Next

Concept Id	Concept Code	Concept Name	Domain	Standard Concept Caption	Exclude	Descendants	Mapped
1118084	140587	celecoxib	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Classification Non-Standard Standard

Delete Concept Set Close Concept Set

- Use of OHDSI standardized vocabularies enables efficient definition of concept sets, which can be fully expressed as all included concepts and included source codes
- Use of standardized vocabularies enables same definition to be applied across different databases, even if those databases use different source coding

# Implementing the target cohort in ATLAS: Review 'celecoxib' included concepts

Concept Set Expression   **Included Concepts** 193   Included Source Codes   Export

▼ Vocabulary  
RxNorm (22)  
DPD (0)  
DA\_France (0)

▼ Class  
Clinical Drug (6)  
Branded Drug Comp (4)  
Branded Drug (4)  
Clinical Drug Comp (4)  
Clinical Drug Form (2)  
Branded Drug Form (1)  
Ingredient (1)  
Branded Drug Rx (0)

▼ Domain  
Drug (22)

▼ Standard Concept  
Standard (22)

▼ Invalid Reason  
Valid (22)

▼ Has Records  
false (13)  
true (9)

▼ Has Descendant Records  
true (19)  
false (3)

Show All entries

Showing 1 to 22 of 22 entries

Filter:

Previous 1 Next

	Id	Code	Name	Class	RC	DRC	Domain	Vocabulary
🛒	1118084	140587	celecoxib	Ingredient	2,399,235	9,052,083	Drug	RxNorm
🛒	40092942	371343	celecoxib Oral Capsule	Clinical Drug Form	0	6,652,848	Drug	RxNorm
🛒	40092943	366443	celecoxib Oral Capsule [Celebrex]	Branded Drug Form	0	6,221,423	Drug	RxNorm
🛒	19029025	205323	celecoxib 200 MG Oral Capsule	Clinical Drug	386,080	6,147,286	Drug	RxNorm
🛒	19081053	315604	celecoxib 200 MG	Clinical Drug Comp	0	6,147,286	Drug	RxNorm
🛒	1118088	213469	celecoxib 200 MG Oral Capsule [Celebrex]	Branded Drug	5,761,206	5,761,206	Drug	RxNorm
🛒	19058155	573357	celecoxib 200 MG [Celebrex]	Branded Drug Comp	0	5,761,206	Drug	RxNorm
🛒	19029024	205322	celecoxib 100 MG Oral Capsule	Clinical Drug	42,124	475,165	Drug	RxNorm
🛒	19081052	315603	celecoxib 100 MG	Clinical Drug Comp	0	475,165	Drug	RxNorm
🛒	1118087	213468	celecoxib 100 MG Oral Capsule [Celebrex]	Branded Drug	433,041	433,041	Drug	RxNorm
🛒	19058154	573356	celecoxib 100 MG [Celebrex]	Branded Drug Comp	0	433,041	Drug	RxNorm
🛒	1118091	349514	celecoxib 400 MG Oral Capsule	Clinical Drug	2,142	26,108	Drug	RxNorm
🛒	19097794	350656	celecoxib 400 MG	Clinical Drug Comp	0	26,108	Drug	RxNorm
🛒	1118113	352314	celecoxib 400 MG Oral Capsule [Celebrex]	Branded Drug	23,966	23,966	Drug	RxNorm
🛒	19121004	576008	celecoxib 400 MG [Celebrex]	Branded Drug Comp	0	23,966	Drug	RxNorm
🛒	1118115	686379	celecoxib 50 MG Oral Capsule	Clinical Drug	1,079	4,289	Drug	RxNorm
🛒	1118114	686378	celecoxib 50 MG	Clinical Drug Comp	0	4,289	Drug	RxNorm
🛒	1118116	686381	celecoxib 50 MG Oral Capsule [Celebrex]	Branded Drug	3,210	3,210	Drug	RxNorm
🛒	19125422	686380	celecoxib 50 MG [Celebrex]	Branded Drug Comp	0	3,210	Drug	RxNorm
🛒	40092944	438305	celecoxib Oral Tablet	Clinical Drug Form	0	0	Drug	RxNorm

- RxNorm is a standard vocabulary to represent drugs
- Descendant concepts from RxNorm ingredient includes clinical drugs, branded drugs, clinical/brand drug forms, and clinical/branded drug component
- RC: 'record count' = how often that standard concept appeared directly in a database
- DRC: 'descendant record count' = how often that standard concept or any of its descendant concepts appeared in a database





# Implementing the target cohort in ATLAS: Review 'celecoxib' included source codes

Concept Set Expression   Included Concepts **193**   **Included Source Codes**   Export

▼ Vocabulary  
NDC (711)  
DPD (125)  
DA\_France (71)  
SPL (34)  
RxNorm (22)  
VA Product (14)  
SNOMED (7)  
GPI (5)

▼ Invalid Reason  
Valid (816)  
Invalid (198)

▼ Class  
11-digit NDC (611)  
9-digit NDC (100)  
Branded Drug (69)  
Branded Drug Comp (66)  
Branded Drug Form (33)  
Prescription Drug (31)  
Clinical Drug (20)  
Branded Drug Box (18)

▼ Domain  
Drug (1006)  
null (7)  
Observation (1)

Show **15** entries

Showing 1 to 15 of 1,014 entries

Filter:

Previous **1** 2 3 4 5 ... 68 Next

	<b>Id</b>	<b>Code</b>	<b>Name</b>	<b>Class</b>	<b>Domain</b>	<b>Vocabulary</b>
🛒	46366487	138110661	<a href="#">celecoxib 400mg/1 ORAL CAPSULE</a>	9-digit NDC	Drug	NDC
🛒	46366486	138110660	<a href="#">celecoxib 200mg/1 ORAL CAPSULE</a>	9-digit NDC	Drug	NDC
🛒	46366485	138110659	<a href="#">celecoxib 100mg/1 ORAL CAPSULE</a>	9-digit NDC	Drug	NDC
🛒	46366484	138110658	<a href="#">celecoxib 50mg/1 ORAL CAPSULE</a>	9-digit NDC	Drug	NDC
🛒	46366319	105440929	<a href="#">celecoxib 200mg/1 ORAL CAPSULE</a>	9-digit NDC	Drug	NDC
🛒	46366233	009046503	<a href="#">celecoxib 200mg/1 ORAL CAPSULE</a>	9-digit NDC	Drug	NDC
🛒	46366232	009046502	<a href="#">celecoxib 100mg/1 ORAL CAPSULE</a>	9-digit NDC	Drug	NDC
🛒	46365423	68180039807	<a href="#">celecoxib 400 MG Oral Capsule</a>	11-digit NDC	Drug	NDC
🛒	46365036	43353003630	<a href="#">celecoxib 200 MG Oral Capsule</a>	11-digit NDC	Drug	NDC
🛒	46332816	13811065830	<a href="#">celecoxib 50 MG Oral Capsule</a>	11-digit NDC	Drug	NDC
🛒	46332722	13811066150	<a href="#">celecoxib 400 MG Oral Capsule</a>	11-digit NDC	Drug	NDC
🛒	46332636	13811065901	<a href="#">celecoxib 100 MG Oral Capsule</a>	11-digit NDC	Drug	NDC
🛒	46332619	13811066001	<a href="#">celecoxib 200 MG Oral Capsule</a>	11-digit NDC	Drug	NDC
🛒	46332520	13811065930	<a href="#">celecoxib 100 MG Oral Capsule</a>	11-digit NDC	Drug	NDC
🛒	46332241	13811065910	<a href="#">celecoxib 100 MG Oral Capsule</a>	11-digit NDC	Drug	NDC

- Many different source vocabularies used across various health systems are mapped into one common reference standard used in OMOP Common Data Model (ex: NDC, DPD, DA France, VA Product, GPI all mapped into RxNorm)
- By defining a concept set as one standard concept and including all descendants, the definition includes 193 different standard concepts and 1,014 different source vocabulary terms.



# Implementing the target cohort in ATLAS:

## Define 'cancer' concept set

**Concept Set Expression** Included Concepts **3567** Included Source Codes Export

Name:  
Broad malignancies excluding skin cancer (incl. primary, secondary)

Show **25** entries Search:

Showing 1 to 3 of 3 entries Previous **1** Next

	Concept Id	Concept Code	Concept Name	Domain	Standard Concept Caption	Exclude	Descendants	Mapped
	4300118	402815007	Squamous cell carcinoma	Condition	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	443392	363346000	Malignant neoplastic disease	Condition	Standard	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	4179980	429114002	Malignant basal cell neoplasm of skin	Condition	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

■ Classification ■ Non-Standard ■ Standard

Delete Concept Set Close Concept Set

- OHDSI standardized vocabularies allow for use of hierarchical structure contained within vocabularies to define large sets of concepts using a small number of concepts
- Example: to define 'all malignancies except skin cancer', we select all 'malignant neoplastic disease' with associated descendants, but exclude all descendants of both 'squamous cell carcinoma' and 'malignant basal cell neoplasm of skin'
- Expansion of this expression defined by 3 concepts manifest as 3,567 distinct standard concepts and 10,810 included source codes





# Hands-on Exercise

Create a cohort definition to replicate the comparator group used in Garbe et al.

1. Go to: <http://www.ohdsi.org/redshift/atlas>
2. Click on 'Define a New Cohort' button
3. Give your cohort a new name (ex. "OHDSI tutorial Garbe comparator replication by Patrick Ryan")
4. On 'Definition' tab, define cohort entry criteria (initial events and all inclusion criteria) and cohort exit criteria
5. Hit 'Save' button beside the cohort definition name
6. Go to 'Generation' tab, and click 'Generate' button beside whichever database(s) you'd like to explore



# Garbe et al. replication: Implementing the outcome in ATLAS

## Definition of UGIC

Upper gastrointestinal complications were defined as hemorrhage, perforation, or obstruction located in the stomach, duodenum, or gastrojejunal part of the GI tract. The following ICD-10 codes included in the subdivisions hemorrhage and perforation were ascertained for the outcome: gastric ulcer (K25), duodenal ulcer (K26), peptic ulcer (K27), gastrojejunal ulcer (K28), hemorrhage of anus and rectum (K62.5), hematemesis (K92.0), melena (K92.1), and GI hemorrhage unspecified (K92.2). High positive predictive values of site- and lesion-specific codes (between 80 and 97 %) and somewhat lower predictive values of non-specific codes (between 57 and 70 %) have been reported for GI ulcers and complications in the ICD coding system in several studies [13, 14].



# Garbe et al. replication: Implementing the outcome cohort in ATLAS

ATLAS Cohort

OHDSI estimation tutorial: Garbe replication: outcome cohort - Upper gastrointestinal comp Save Close Copy

Definition Concept Sets Generation Reporting Explore Export

**Cohort definition:** A cohort is defined as the set of persons satisfying one or more inclusion criteria for a duration of time. One person may qualify for one cohort multiple times during non-overlapping time intervals. Cohorts are constructed in ATLAS by specifying cohort entry criteria and cohort exit criteria. Cohort entry criteria involve selecting one or more initial events, which determine the start date for cohort entry, and optionally specifying additional inclusion criteria which filter to the qualifying events. Cohort exit criteria are applied to each cohort entry record to determine the end date when the person's episode no longer qualifies for the cohort.

Upper gastrointestinal complication (UGIC) events, as defined in Garbe et al, Eur J Clin Pharmacol, 2012; <http://www.ncbi.nlm.nih.gov/pubmed/22763756>

All Cohort Entry Criteria Cohort Exit Criteria

**Initial event cohort:** Events are recorded time-stamped observations for the persons, such as drug exposures, conditions, procedures, measurements and visits. All events have a start date and end date, though some events may have a start date and end date with the same value (such as procedures or measurements). The event index date is set to be equal to the event start date.

People having any of the following: Add Initial Event...

a condition era of Upper gastrointestinal complication (UGIC) Add Add criteria attribute... Delete Criteria

with continuous observation of at least 0 days before and 0 days after event index date

Limit initial events to: all events per person.

**Index event inclusion criteria:** From among the index events, include:

People having all of the following criteria: Add New Criteria...

with exactly 0 using all occurrences of:

a condition occurrence of Upper gastrointestinal complication (UGIC) Add Add criteria attribute...

occurring between 30 days Before and 1 days Before event index date Delete Criteria

Ambiguities from publication which require detailed specification in a complete protocol:

- 1) How do we determine distinct events (and not misclassification continuation of care for prior episode as incident occurrence)?
- 2) How does 'validation' of ICD9 codes in Italy and Canada improve your confidence in accuracy of ICD10 codes in Germany?



# Implementing the outcome cohort in ATLAS:

## Define 'UGIC' concept set

[Concept Set Expression](#) [Included Concepts 325](#) [Included Source Codes](#) [Export](#)

Name:

Show  entries Search:

Showing 1 to 14 of 14 entries Previous  Next

	Concept Id	Concept Code	Concept Name	Domain	Standard Concept Caption	Exclude	Descendants	Mapped
	46273478	1092881000119105	Rectal hemorrhage due to ulcerative colitis	Condition	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	194158	48729005	Perinatal gastrointestinal hemorrhage	Condition	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	4172869	276525003	Peptic ulcer of newborn	Condition	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	4027663	13200003	Peptic ulcer	Condition	Standard	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	4048286	206425006	Neonatal rectal hemorrhage	Condition	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	4048608	206424005	Neonatal melena	Condition	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	4071070	206423004	Neonatal hematemesis	Condition	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	4103703	2901004	Melena	Condition	Standard	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	316457	35265002	Mallory-Weiss syndrome	Condition	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	4059178	16121001	Gastrojejunal ulcer	Condition	Standard	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	192671	74474003	Gastrointestinal hemorrhage	Condition	Standard	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	4265600	397825006	Gastric ulcer	Condition	Standard	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	4198381	51868009	Duodenal ulcer disease	Condition	Standard	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	26441	57748001	Bleeding ulcer of esophagus	Condition	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Classification

Non-Standard

Standard

[Delete Concept Set](#) [Close Concept Set](#)

- Standard concept set can be constructed that yields a specific set of source codes
- Standard concepts can then be applied to other databases that use different source codes



# Implementing the outcome cohort in ATLAS:

## Define 'UGIC' concept set

Concept Set Expression   Included Concepts 325   Included Source Codes   Export

Vocabulary

- ICD10CM (47)
- SNOMED (0)
- CIEL (0)
- Read (0)
- ICD9CM (0)
- ICD10 (0)
- OXMIS (0)
- MeSH (0)

Invalid Reason

- Valid (47)
- Invalid (0)

Class

- 4-char billing code (40)
- 3-char nonbill code (4)
- 5-char billing code (3)
- 4-dig nonbill code (0)
- OXMIS (0)
- Diagnosis (0)
- 5-dig billing code (0)
- 4-dig billing code (0)

Domain

- Condition (47)
- null (0)
- Observation (0)

Show All entries

Showing 1 to 47 of 47 entries

	Id	Code	Name	Class	Domain	Vocabulary
	35208201	K25.0	Acute gastric ulcer with hemorrhage	4-char billing code	Condition	ICD10CM
	1569562	K25	Gastric ulcer	3-char nonbill code	Condition	ICD10CM
	35208202	K25.1	Acute gastric ulcer with perforation	4-char billing code	Condition	ICD10CM
	35208203	K25.2	Acute gastric ulcer with both hemorrhage and perforation	4-char billing code	Condition	ICD10CM
	35208204	K25.3	Acute gastric ulcer without hemorrhage or perforation	4-char billing code	Condition	ICD10CM
	35208205	K25.4	Chronic or unspecified gastric ulcer with hemorrhage	4-char billing code	Condition	ICD10CM
	35208206	K25.5	Chronic or unspecified gastric ulcer with perforation	4-char billing code	Condition	ICD10CM
	35208207	K25.6	Chronic or unspecified gastric ulcer with both hemorrhage and perforation	4-char billing code	Condition	ICD10CM
	35208208	K25.7	Chronic gastric ulcer without hemorrhage or perforation	4-char billing code	Condition	ICD10CM
	35208209	K25.9	Gastric ulcer, unspecified as acute or chronic, without hemorrhage or perforation	4-char billing code	Condition	ICD10CM
	35208210	K26.0	Acute duodenal ulcer with hemorrhage	4-char billing code	Condition	ICD10CM
	1569563	K26	Duodenal ulcer	3-char nonbill code	Condition	ICD10CM
	35208211	K26.1	Acute duodenal ulcer with perforation	4-char billing code	Condition	ICD10CM
	35208212	K26.2	Acute duodenal ulcer with both hemorrhage and perforation	4-char billing code	Condition	ICD10CM

- 47 distinct ICD10CM codes map to standard concepts
- Complete listing required for full transparency, rather than assuming user knows subcodes within hierarchy (e.g. NEVER WRITE ICD9 ###.\*)

perforation



# Implementing the outcome cohort in ATLAS: Define 'UGIC' concept set

Concept Set Expression   Included Concepts 325   Included Source Codes   Export

▼ Vocabulary  
SNOMED (478)  
CIEL (191)  
Read (157)  
ICD9CM (126)  
ICD10CM (47)  
ICD10 (45)  
OXMIS (27)  
MeSH (9)

▼ Invalid Reason  
Valid (923)  
Invalid (158)

▼ Class  
Clinical Finding (477)  
Diagnosis (191)  
Read (157)  
5-digit billing code (82)  
ICD10 code (41)  
4-char billing code (40)  
4-digit nonbill code (36)  
OXMIS (27)

▼ Domain  
Condition (1079)  
null (1)  
Observation (1)

Show  entries

Showing 1 to 1,081 of 1,081 entries

Filter:

Previous  Next

	Id	Code	Name	Class	Domain	Vocabulary
🛒	45757062	103691000119106	Gastric ulcer due to Helicobacter pylori	Clinical Finding	Condition	SNOMED
🛒	432951	10389003	Acute gastrojejunal ulcer without hemorrhage AND without perforation but with obstruction	Clinical Finding	Condition	SNOMED
🛒	4027942	10699001	Esophagogastric ulcer	Clinical Finding	Condition	SNOMED
🛒	46269819	1082631000119106	Gastric hemorrhage due to allergic gastritis	Clinical Finding	Condition	SNOMED
🛒	46269820	1082671000119109	Haemorrhage of anastomosis due to ulcer	Clinical Finding	Condition	SNOMED
🛒	46269821	1082701000119105	Duodenal hemorrhage due to angiodysplasia of duodenum	Clinical Finding	Condition	SNOMED
🛒	46269822	1082711000119108	Gastric hemorrhage due to angiodysplasia of stomach	Clinical Finding	Condition	SNOMED
🛒	46269837	1085121000119107	Gastric hemorrhage due to chronic superficial gastritis	Clinical Finding	Condition	SNOMED
🛒	46269847	1085221000119103	Rectal hemorrhage due to chronic ulcerative proctitis	Clinical Finding	Condition	SNOMED
🛒	46269863	1085431000119101	Rectal hemorrhage due to inflammatory polyps of colon	Clinical Finding	Condition	SNOMED
🛒	46269877	1085791000119105	Rectal hemorrhage due to Crohn's disease of large intestine	Clinical Finding	Condition	SNOMED
🛒	46269883	1085841000119108	Rectal hemorrhage due to Crohn's disease of small intestine	Clinical Finding	Condition	SNOMED

- 47 distinct ICD10CM codes map to standard concepts...
- ...but so do 126 ICD9CM codes, 157 Read codes, 27 OXMIS codes, etc.
- Using one standard concept definition allows consistent application of clinical construct across different databases, even if they use different source vocabularies
- Cross-database analyses require review of standard concepts and mapped source codes





# Garbe et al. replication: Designing the statistical analysis in ATLAS

## Statistical analysis

We used a Poisson regression analysis to estimate the rate ratio (RR) of UGIC for coxib initiation versus tNSAID initiation and its 95 % confidence interval (CI). A conventional approach and the hd-PS approach were used to estimate the PS. The PS was estimated as the probability of initiating a coxib in a logistic regression model. In the conventional approach, the PS was estimated via a logistic regression model for coxib initiation that included all 79 pre-specified covariates described above, which were ascertained during the 6-month period before cohort entry.



# Garbe et al. replication: Designing the statistical analysis in ATLAS

ATLAS x

hix.jnj.com/atlas/#/estimation/5

**ATLAS**

- Home
- Data Sources
- Vocabulary
- Concept Sets
- Cohorts
- Profiles
- Estimation
- Jobs
- Configuration
- Feedback

## Population Level Effect Estimation

OHDSI estimation tutorial: Garbe replication: celecoxib vs. diclofenac for rate of upper gastro Save Close

Specification Export

Choose your target cohort:

OHDSI estimation tutorial: Garbe replication: target cohort - celecoxib new users 📄

Choose your comparator cohort:

OHDSI estimation tutorial: Garbe replication: comparator cohort - diclofenac new users 📄

Choose your outcome cohort:

OHDSI estimation tutorial: Garbe replication: outcome cohort - Upper gastrointestinal complication (UGIC) events 📄

Specify the statistical model used to estimate the risk of outcome between target and comparator cohorts:

Poisson regression ▼

Define the time-at-risk window start, relative to target/comparator cohort entry:

0 ▼ days from cohort start date

Define the time-at-risk window end:

0 ▼ days from cohort end date ▼

Minimum washout period applied to target and comparator cohorts:

0 ▼

Minimum required days at risk, applied to target and comparator cohorts:

0 ▼

Remove patients who enter both cohorts?

Yes ▼

Remove patients who have observed the outcome prior to cohort entry?

No ▼

Use propensity score adjustment as a confounding adjustment strategy for baseline covariates?

Yes ▼





# The choice of the outcome model defines your research question

	Logistic regression	Poisson regression	Cox proportional hazards
How the outcome cohort is used	Binary classifier of presence/absence of outcome during the fixed time-at-risk period	Count the number of occurrences of outcomes during time-at-risk,	Compute time-to-event from time-at-risk start until earliest of first occurrence of outcome or time-at-risk end, and track the censoring event (outcome or no outcome)
‘Risk’ metric	Odds ratio	Rate ratio	Hazards ratio
Key model assumptions	Constant response in fixed window	Outcomes follow Poisson distribution	Proportionality – constant relative hazard



# Cohort restriction decisions

Specify the statistical model used to estimate the risk of outcome between target and comparator cohorts:

Poisson regression ▼

Define the time-at-risk window start, relative to target/comparator cohort entry:

0 ▼ days from cohort start date

Define the time-at-risk window end:

0 ▼ days from cohort end date ▼

Minimum washout period applied to target and comparator cohorts:

0 ▼

Minimum required days at risk, applied to target and comparator cohorts:

0 ▼

Remove patients who enter both cohorts?

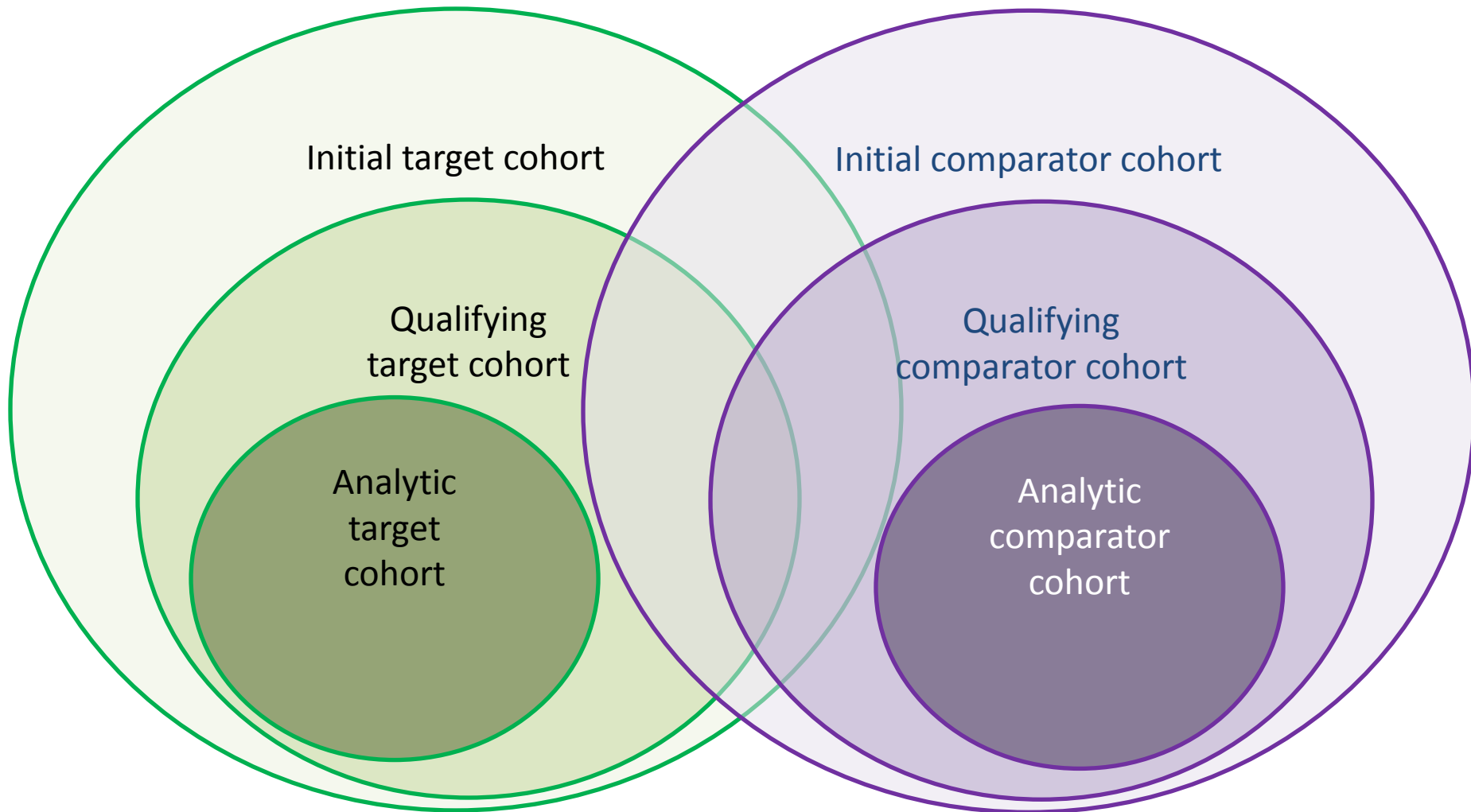
Yes ▼

Remove patients who have observed the outcome prior to cohort entry?

No ▼



# Cohort restriction in comparative cohort analyses





# Two forms of attrition to consider as diagnostics

1. Initial cohort → Qualifying cohort: (independent from analysis)

How did additional inclusion criteria impact the proportion and composition of your cohort?

Graham replication: comparator cohort – warfarin new users

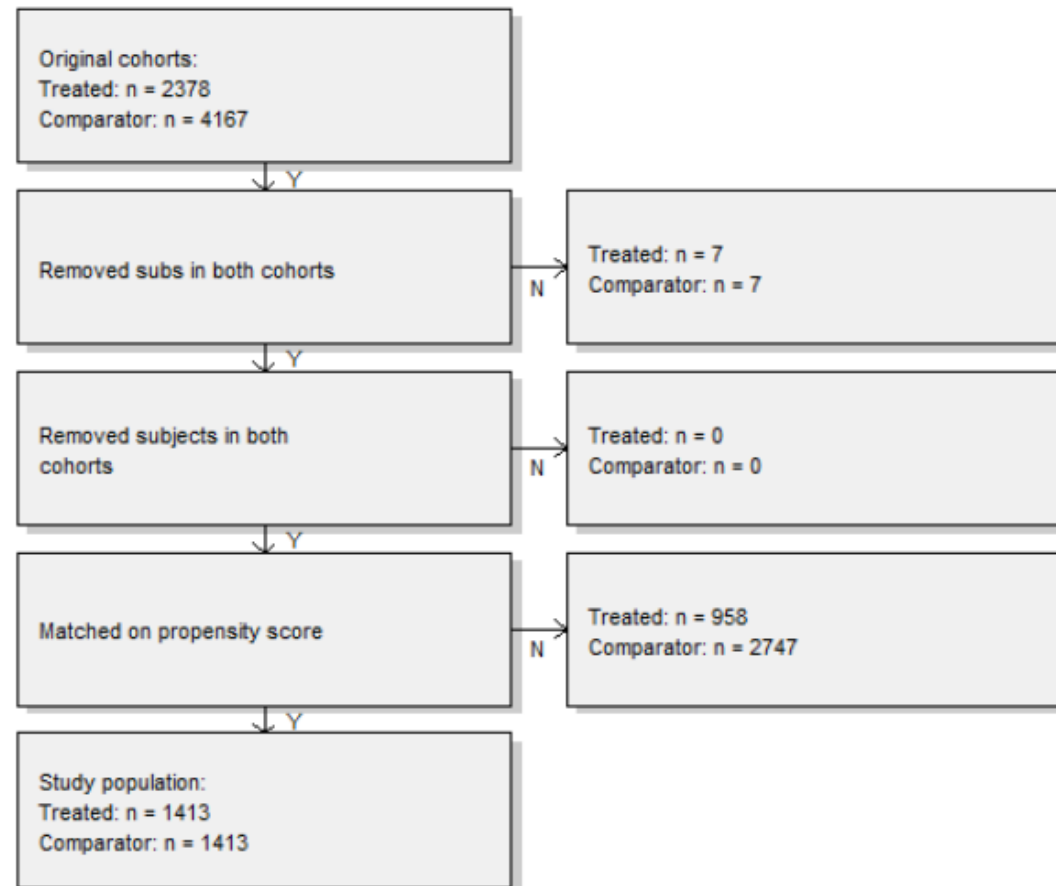
		Match Rate	Matches	Total
Summary Statistics:		31.52%	52,400	166,243
Inclusion Rule		N	% Remain	% Diff
1. Has prior atrial fibrillation or atrial flutter diagnosis		78,371	47.14%	52.86%
2. Has no prior treatment with comparator drug (dabigatran)		74,931	45.07%	2.07%
3. Has no prior treatment with other anticoagulants (rivaroxaban or apixaban)		71,879	43.24%	1.84%
4. Not in a skilled nursing facility or nursing home, or receiving hospice care on the index date		71,834	43.21%	0.03%
5. Not undergoing dialysis or kidney transplant recipient		70,148	42.20%	1.01%
6. No mitral valve disease, heart valve repair, or replacement in the prior 6 months		64,580	38.85%	3.35%
7. No deep vein thrombosis or pulmonary embolism in the prior 6 months		54,791	32.96%	5.89%
8. No joint replacement surgery in the prior 6 months		52,400	31.52%	1.44%



# Two forms of attrition to consider as diagnostics

## 2. Qualifying cohort → Analytic cohort

How did analysis restrictions impact the proportion and composition of your cohort?





# Covariate adjustment strategy

Use propensity score adjustment as a confounding adjustment strategy for baseline covariates?

No ▼

Do you want to adjust for baseline covariates in the outcome model?

No ▼

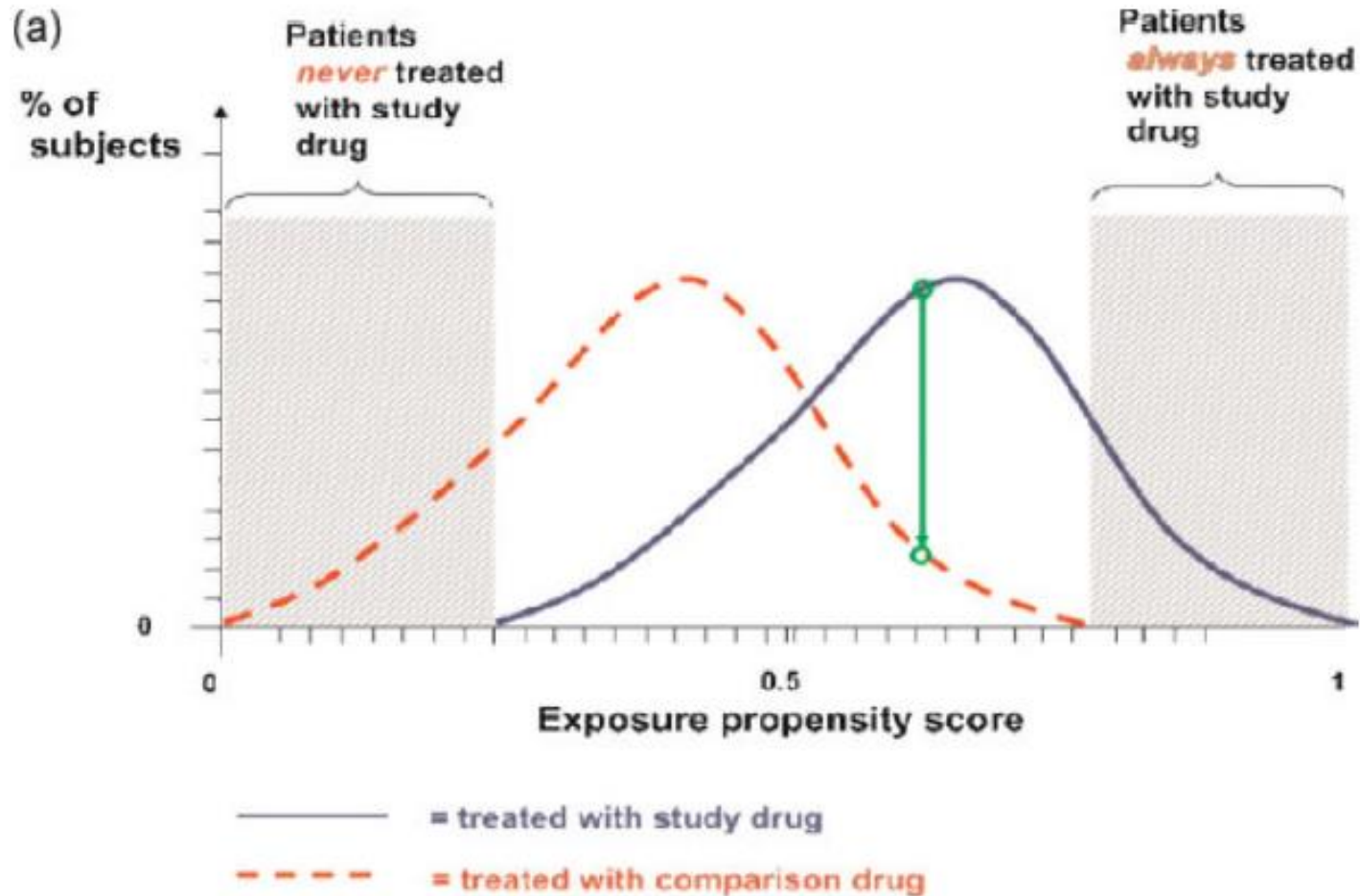


# Propensity score introduction

- $e(x) = \Pr(Z=1 | x)$ 
  - $Z$  is treatment assignment
  - $x$  is a set of all covariates at the time of treatment assignment
- Propensity score = probability of belonging to the target cohort vs. the comparator cohort, given the baseline covariates
- Propensity score can be used as a ‘balancing score’: if the two cohorts have similar propensity score distribution, then the distribution of covariates should be the similar (need to perform diagnostic to check)



# Intuition around propensity score balance







# “Five reasons to use propensity score in pharmacoepidemiology”

- Theoretical advantages
  - Confounding by indication is the primary threat to validity, PS focuses directly on indications for use and non-use of drug under study
- Value of propensity scores for matching or trimming the population
  - Eliminate ‘uncomparable’ controls without assumptions of linear relationship between PS and outcome
- Improved estimation with few outcomes
  - PS allows matching on one scalar value rather than needing degrees of freedom for all covariates
- Propensity score by treatment interactions
  - PS enables exploration of patient-level heterogeneity in response
- Propensity score calibration to correct for measurement error



# Covariate selection in propensity score modeling

- What covariates should you include in propensity score model?
  - Variables that predict exposure status (Rubin Biometrika 1983)
  - Variables that are confounders, associated with both exposure and outcome (Schneeweiss Epidemiology 2009)
  - Variables that are associated with outcomes (Brookhart AJE 2006)
- Propensity score tends to balance distributions of covariates used in estimation
  - The method does NOTHING for unmeasured confounding or other covariates not entered into model



# My perspective on covariate selection

- Choosing the ‘right’ variables in the model is an empirical question. It is the set of variables that yield the unbiased estimate of the effect of interest.
- The goal of fitting a propensity score is to predict treatment assignment, so a reasonable objective function is to maximize discrimination (AUC)
- Large-scale regression, using L1 regularization (LASSO), that uses a large set of potential covariates will often outperform a traditional regression that uses a small subset of those covariates
  - Regularization reduces risk of model overfitting, by only selecting the covariates that have an adequate information component
  - Covariates that aren’t used are effectively ‘unmeasured’



# Covariate selection in ATLAS

Use propensity score adjustment as a confounding adjustment strategy for baseline covariates?

Yes ▼

Which types of baseline covariates do you want to include in the propensity score model?

- ☒ Demographics
  - ☒ Gender
  - ☒ Age group (5-year bands)
  - ☒ Index year
  - ☐ Index month
  - ☐ Race
  - ☐ Ethnicity
- ☒ Conditions
  - ☐ In prior 30d
  - ☒ In prior 365d
  - ☐ In prior 180d within inpatient setting
  - ☐ All time prior
  - ☐ Overlapping index date
- ☐ Condition aggregation
  - ☐ SNOMED
  - ☐ MedDRA
- ☒ Drugs
  - ☐ In prior 30d
  - ☐ In prior 365d
  - ☐ All time prior
  - ☒ Overlapping index date
- ☒ Drug aggregation
  - ☐ Clinical Drug
  - ☒ Ingredient
  - ☒ ATC Class
- ☐ Procedures
  - ☐ In prior 30d
  - ☐ In prior 365d
- ☐ Measurement
  - ☐ Existence in prior 30d
  - ☐ Existence in prior 365d
  - ☐ Count in prior 365d
  - ☐ Has latest prior numeric value below normal range
  - ☐ Has latest prior numeric value above normal range
- ☐ Risk scores
  - ☐ Charlson
  - ☐ CHADS2
  - ☐ DCSI
- ☒ Concept counts (count of distinct conditions/procedures/visits in history)
- ☐ Interaction terms
  - ☐ By index year
  - ☐ By index month

What concepts do you want to include in baseline covariates in the propensity score model? (Leave blank if you want to include everything)

OHDSI estimation tutorial- Garbe replication: covariates to include in PS model



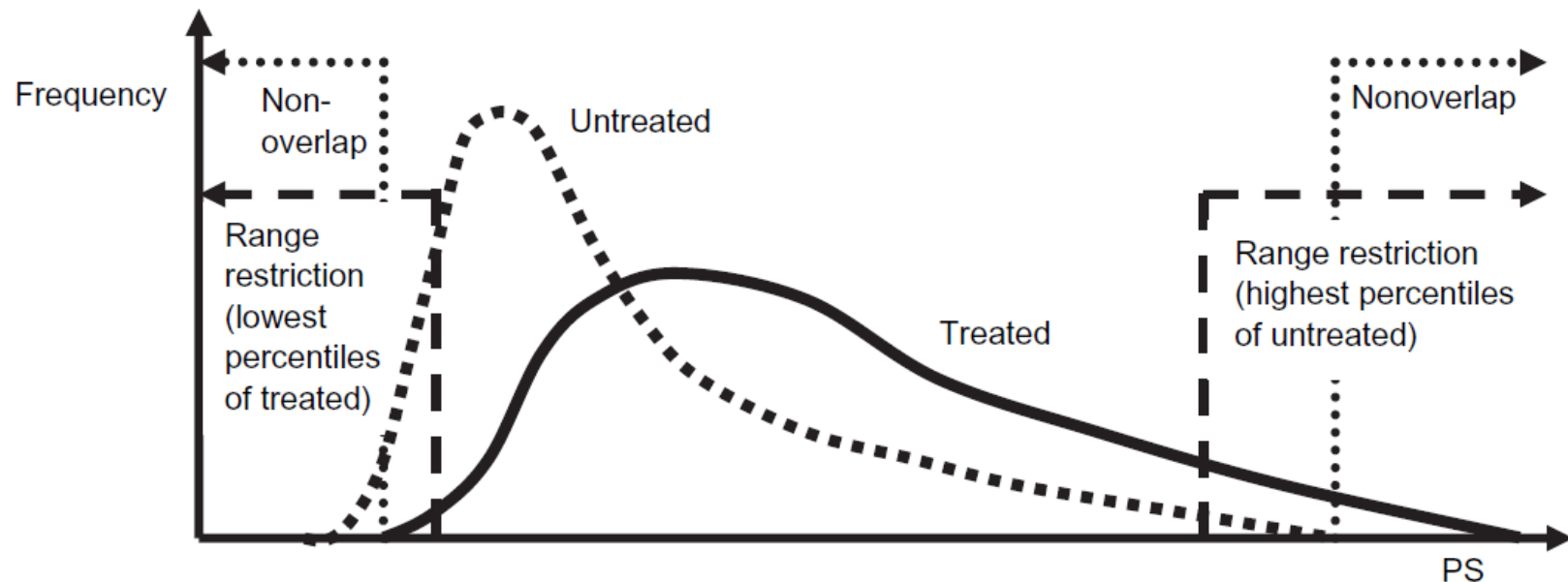
What concepts do you want to exclude from baseline covariates in the propensity score model? (Leave blank if you want to include everything)

OHDSI estimation tutorial- Garbe replication: covariates to exclude in PS model





# Design choice: propensity score trimming by percentile



- Simulation studies suggest PS trimming may eliminate confounding due to extreme patients with 'last resort treatment' or 'treatment withhold'
- The subpopulation you select may be systematically different from the overall population



# Propensity score trimming by percentile in ATLAS

How do you want to restrict your cohorts based on the propensity score distribution?

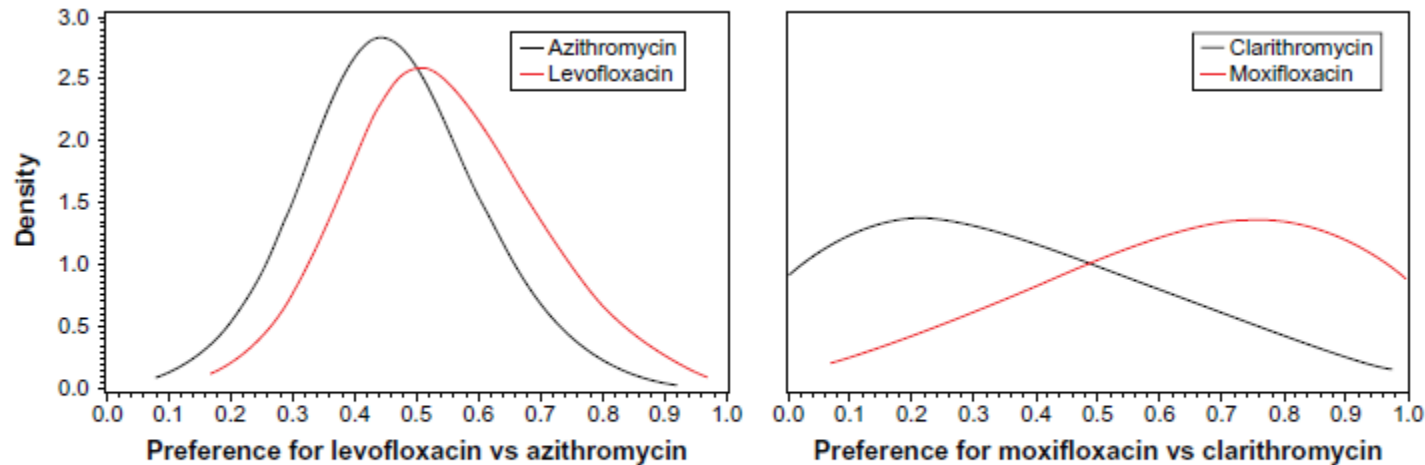
by Percentile ▼

Trim Fraction (1-100%):

5



# Design choice: propensity score trimming by equipoise



**Figure 1** Preference score distributions.

**Notes:** Preference distributions for a pair of antibiotics given to very similar patients (left) and for a pair given to substantially different patient populations. (See Table 1 for salient differences).

- Preference score (PREF) = propensity score, weighted for the imbalance in the prevalence of the target vs. comparator cohort
- PREF= 0.5 means equally likely to belong to either cohort
- Trimming to PREF near 0.5 restricts to persons who had reasonable probability of assignment in both groups ('near clinical equipoise')



# Propensity score trimming by equipoise in ATLAS

How do you want to restrict your cohorts based on the propensity score distribution?

by Equipoise ▼

Trim Fraction (1-100%):

5





# Methods for confounding adjustment using a propensity score

Regression adjustment	The PS is used as a covariable in an outcome regression model to adjust the as assum same relationship between propensity score and outcome is correctly specified.
Matching	The PS is used to match exposed subjects to unexposed subjects with similar values of the PS. This method assumes that within the matched sample, exposed and unexposed subjects have a similar distribution of baseline characteristics.
Stratification	The PS is used to stratify subjects into (often quintiles or deciles) strata. Treatment effects are estimated separately within each stratum and then combined into an overall estimate of treatment effect. This method assumes that within each stratum, exposed and unexposed subjects have a similar distribution of baseline characteristics.
Inverse Probability Weighting	The PS is used to create weights based on the inverse probability which is defined as: $E^*/PS + (1-E)/(1-PS)$ . This assumes that baseline characteristics are similar in the exposed and unexposed group.

Not generally recommended

Fully implemented in OHDSI CohortMethod R package

\* E: exposure



# Propensity score adjustment in ATLAS

## Matching:

Do you want to perform matching or stratification?

Matching ▼

How many comparator patients do you want to select for each target patient (within a defined caliper)?

1

## Stratification:

Do you want to perform matching or stratification?

Stratification ▼

How many strata do you want to use?

5



# Outcome model covariate adjustment

- Final outcome model can be univariate (estimate effect of cohort class on outcome alone) or multivariate (estimate effect of cohort class on outcome, adjusting for other baseline covariates)
- If propensity score matching or stratification is used, outcome model should be conditional regression (estimate effect of cohort class on outcome within each matched set)
- Outcome model typically bounded by degrees of freedom; can only include additional covariates if sufficient number of outcomes (rule of thumb: 10 outcomes per extra covariate)



# Outcome model covariate adjustment in ATLAS

Do you want to adjust for baseline covariates in the outcome model?

Yes ▼

Which types of baseline covariates do you want to include in the outcome model?

- ☐ Demographics
  - ☐ Gender
  - ☐ Age group (5-year bands)
  - ☐ Index year
  - ☐ Index month
  - ☐ Race
  - ☐ Ethnicity
- ☐ Conditions
  - ☐ In prior 30d
  - ☐ In prior 365d
  - ☐ In prior 180d within inpatient setting
  - ☐ All time prior
  - ☐ Overlapping index date
- ☐ Condition aggregation
  - ☐ SNOMED
  - ☐ MedDRA
- ☐ Drugs
  - ☐ In prior 30d
  - ☐ In prior 365d
  - ☐ All time prior
  - ☐ Overlapping index date
- ☐ Drug aggregation
  - ☐ Clinical Drug
  - ☐ Ingredient
  - ☐ ATC Class
- ☐ Procedures
  - ☐ In prior 30d
  - ☐ In prior 365d
- ☐ Measurement
  - ☐ Existence in prior 30d
  - ☐ Existence in prior 365d
  - ☐ Count in prior 365d
  - ☐ Has latest prior numeric value below normal range
  - ☐ Has latest prior numeric value above normal range
- ☐ Risk scores
  - ☐ Charlson
  - ☐ CHADS2
  - ☐ DCSI
- ☐ Concept counts (count of distinct conditions/procedures/visits in history)
- ☐ Interaction terms
  - ☐ By index year
  - ☐ By index month

What concepts do you want to include in baseline covariates in the outcome model? (Leave blank if you want to include everything)



What concepts do you want to exclude from baseline covariates in the outcome model? (Leave blank if you want to include everything)





# Negative control outcomes for empirical calibration

- Observational data analyses may have residual bias, so it's important to perform diagnostics to quantify the extent of this potential issue
- Bias = expected value of the error distribution (random + systematic)
- Negative control outcomes can be used efficiently in cohort analyses
  - Outcomes which have no evidence about association with either target cohort or outcome cohort, therefore 'true RR' assumed to equal 1 and any difference between effect estimate and 'true RR' can be classified as systematic error
  - Convention: find outcomes where 'absence of evidence' can be inferred to be 'evidence of absence':
    1. not listed on target/comparator product labels
    2. not co-occurring with target/comparators in published literature (Medline)
    3. don't have increased signal score from spontaneous adverse event reporting (FAERS)
    4. do appear with adequate prevalence in the observational database so that an effect could have been previously observable had it existed
- Sample of negative control outcomes ( $n > 20$ ) can be used to estimate 'empirical null' distribution, which can then be used to empirically calibrate p-value for unknown outcome of interest



# Pleasure reading to motivate use of negative controls

Statistics  
in Medicine

## Research Article

Received 12 November 2012,

Accepted 3 July 2013

Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5925

## Interpreting observational studies: why empirical calibration is needed to correct $p$ -values

**Martijn J. Schuemie**<sup>a,b,\*†</sup> **Patrick B. Ryan**<sup>b,c</sup>  
**William DuMouchel**<sup>b,d</sup> **Marc A. Suchard**<sup>b,e</sup> and **David Madigan**<sup>b,f</sup>

# Negative control selection in ATLAS

What outcomes would you like to use as your negative controls? These are concepts known not to be associated with either the target or comparator group, such that we can assume the true relative risk should equal 1. These negative control outcomes will be used for empirical calibration.

OHDSI estimation tutorial: Garbe et al. replication, negative controls

## Concept Set

OHDSI estimation tutorial: Garbe et al. replication, negative controls

Save

Close

Concept Set Expression

Included Concepts 34

Included Source Codes

Export

### Vocabulary

SNOMED (34)

### Class

Clinical Finding (34)

### Domain

Condition (34)

### Standard Concept

Standard (34)

### Invalid Reason

Valid (34)

### Has Records

true (34)

### Has Descendant Records

true (34)

Show All entries

Filter:

Previous 1 Next

Showing 1 to 34 of 34 entries

	Id	Code	Name	Class	RC	DRC	Domain	Vocabulary
	257007	61582004	Allergic rhinitis	Clinical Finding	35,111,076	55,582,634	Condition	SNOMED
	255573	13645005	Chronic obstructive lung disease	Clinical Finding	22,235,402	28,534,157	Condition	SNOMED
	319843	11851006	Mitral valve disorder	Clinical Finding	4,805,521	6,232,915	Condition	SNOMED
	141932	398838000	Seborrheic keratosis	Clinical Finding	4,758,455	4,758,455	Condition	SNOMED
	380094	57406009	Carpal tunnel syndrome	Clinical Finding	4,713,600	4,713,600	Condition	SNOMED
	313459	73430006	Sleep apnea	Clinical Finding	4,280,415	17,180,747	Condition	SNOMED
	436665	13746004	Bipolar disorder	Clinical Finding	3,472,356	7,920,853	Condition	SNOMED
	443800	14302001	Amenorrhea	Clinical Finding	3,221,577	3,257,684	Condition	SNOMED
	197236	95315005	Uterine leiomyoma	Clinical Finding	3,021,864	3,898,566	Condition	SNOMED
	321596	20696009	Peripheral venous insufficiency	Clinical Finding	2,950,514	3,004,234	Condition	SNOMED
	436676	47505003	Posttraumatic stress disorder	Clinical Finding	2,933,195	2,978,063	Condition	SNOMED
	139099	400097005	Ingrowing nail	Clinical Finding	2,928,631	2,928,631	Condition	SNOMED
	314658	8186001	Cardiomegaly	Clinical Finding	2,651,255	2,651,255	Condition	SNOMED
	433753	15167005	Alcohol abuse	Clinical Finding	1,758,133	2,013,798	Condition	SNOMED
	440374	191736004	Obsessive-compulsive disorder	Clinical Finding	1,666,847	1,666,981	Condition	SNOMED



# Putting it all together...







# ATLAS print friendly – the start of your team's protocol

## Population Level Effect Estimation

OHDSI estimation tutorial: Garbe replication: celecoxib vs. diclofenac for rate of upper gastro

Save

Close

Specification

Export

Print Friendly

R Code

## Research question

To compare the risk of **OHDSI estimation tutorial: Garbe replication: outcome cohort - Upper gastrointestinal complication (UGIC) events** between **OHDSI estimation tutorial: Garbe replication: target cohort - celecoxib new users** and **OHDSI estimation tutorial: Garbe replication: comparator cohort - diclofenac new users**, we will estimate the population-level effect of exposure on the rate of the outcome during the period from 0 days from cohort start date to 0 days from cohort end date.

## Study Design:

This study will follow a retrospective, observational, comparative cohort design. We define 'retrospective' to mean the study will be conducted using data already collected prior to the start of the study. We define 'observational' to mean there is no intervention or treatment assignment imposed by the study. We define 'cohort' to mean a set of patients satisfying a one or more inclusion criteria for a duration of time. We define 'comparative cohort design' to mean the formal comparison between two cohorts, a target cohort and comparator cohort, for the risk of an outcome during a defined time period after cohort entry.

In this study, we compare **OHDSI estimation tutorial: Garbe replication: target cohort - celecoxib new users** with **OHDSI estimation tutorial: Garbe replication: comparator cohort - diclofenac new users** for the rate of **OHDSI estimation tutorial: Garbe replication: outcome cohort - Upper gastrointestinal complication (UGIC) events** from 0 days from cohort start date to 0 days from cohort end date.

The overall study population could be considered to be patients who entered either the target cohort or comparator cohort. Patients were excluded from consideration if they qualified for both the target cohort and comparator cohort at any time in their record.

The rate of outcomes among patients in the target and comparator cohorts is determined by counting the number of outcome occurrences of **OHDSI estimation tutorial: Garbe replication: outcome cohort - Upper gastrointestinal complication (UGIC) events** during the time-at-risk of 0 days from cohort start date to 0 days from cohort end date.

Propensity scores will be used as an analytic strategy to reduce potential confounding due to imbalance between the target and comparator cohorts in baseline covariates. The propensity score is the probability of a patient being classified in the target cohort vs. the comparator cohort, given a set of observed covariates. In this study, the propensity score is estimated for each patient, using the predicted probability from a regularized logistic regression model, fit with a Laplace prior (LASSO) and the regularization hyperparameter selected by optimizing the likelihood in a 10-fold cross validation, using a starting variance of 0.01 and a tolerance of 2e-7.

The types of baseline covariates used to fit the propensity score model will be:

- Demographics
  - Gender
  - Age group (5-year bands)
  - Index year
- Conditions
  - In prior 365d



# ATLAS R code – the start of your team's implementation

## Population Level Effect Estimation

OHDSI estimation tutorial: Garbe replication: celecoxib vs. diclofenac for rate of upper gastro

Save

Close

Specification

Export

Print Friendly

R Code

```
#####
# Study: OHDSI estimation tutorial: Garbe replication: celecoxib vs. diclofenac for rate of upper gastrointestinal complications (UGIC)
#####

#####
# Cohort Method Installation & Load
#####

# Uncomment to install Cohort Method
# install.packages("drat")
# drat::addRepo(c("OHDSI", "cloudyr"))
# install.packages("CohortMethod")

# Load the Cohort Method library
library(CohortMethod)
library(SqlRender)

#####
# Data extraction
#####

# TODO: Insert your connection details here
connectionDetails <- createConnectionDetails(dbms = "postgresql",
                                             server = "localhost/ohdsi",
                                             user = "joe",
                                             password = "supersecret")

cdmDatabaseSchema <- "my_cdm_data"
resultsDatabaseSchema <- "my_results"
exposureTable <- "exposure_table"
outcomeTable <- "outcome_table"
cdmVersion <- "5"
```



# Hands-on exercise

- Design your study!
  - What's your target cohort?
  - What's your compactor cohort?
  - What's your outcome cohort?
  - What's your time-at-risk?
  - What's your model specification?
  - What's your covariate adjustment strategy?
- Save your progress in ATLAS