OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Creating reproducible studies

Martijn Schuemie

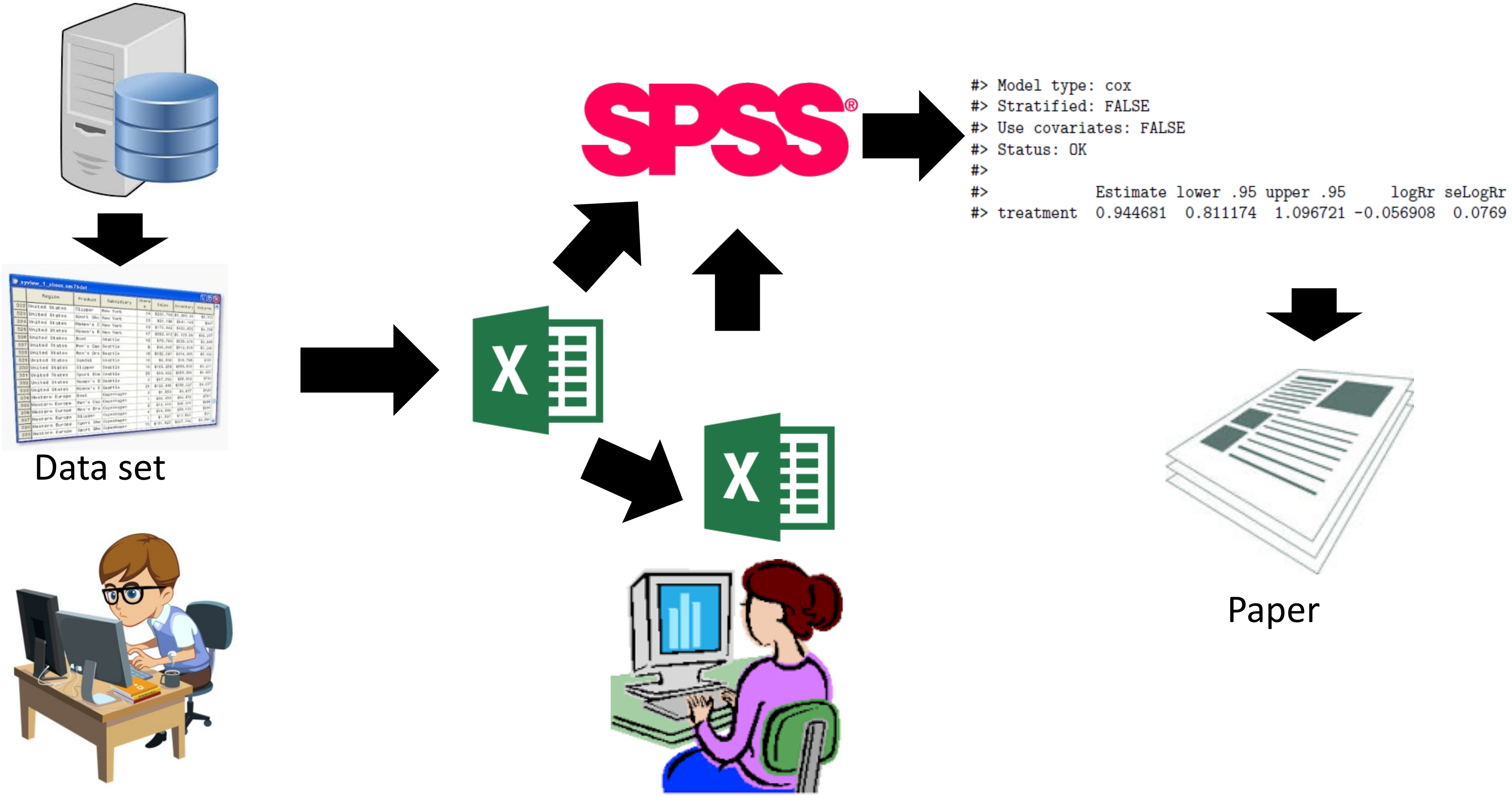
What is reproducibility?

Desired attribute	Question	Researcher	Data	Analysis		Result
Repeatable	Identical	Identical	Identical	Identical	=	Identical
Reproducible	Identical	Different	Identical	Identical	=	Identical
Replicable	Identical	Same or different	Similar	Identical	=	Similar
Generalizable	Identical	Same or different	Different	Identical	=	Similar
Robust	Identical	Same or different	Same or different	Different	=	Similar
Calibrated	Similar (controls)	Identical	Identical	Identical	=	Statistically consistent

Ensuring the **analysis** can be kept identical allows for **repeatable**, **reproducible**, **replicable**, **generalizable**, and **calibrated** science

Source: The Book of OHDSI

What do observational studies currently look like?

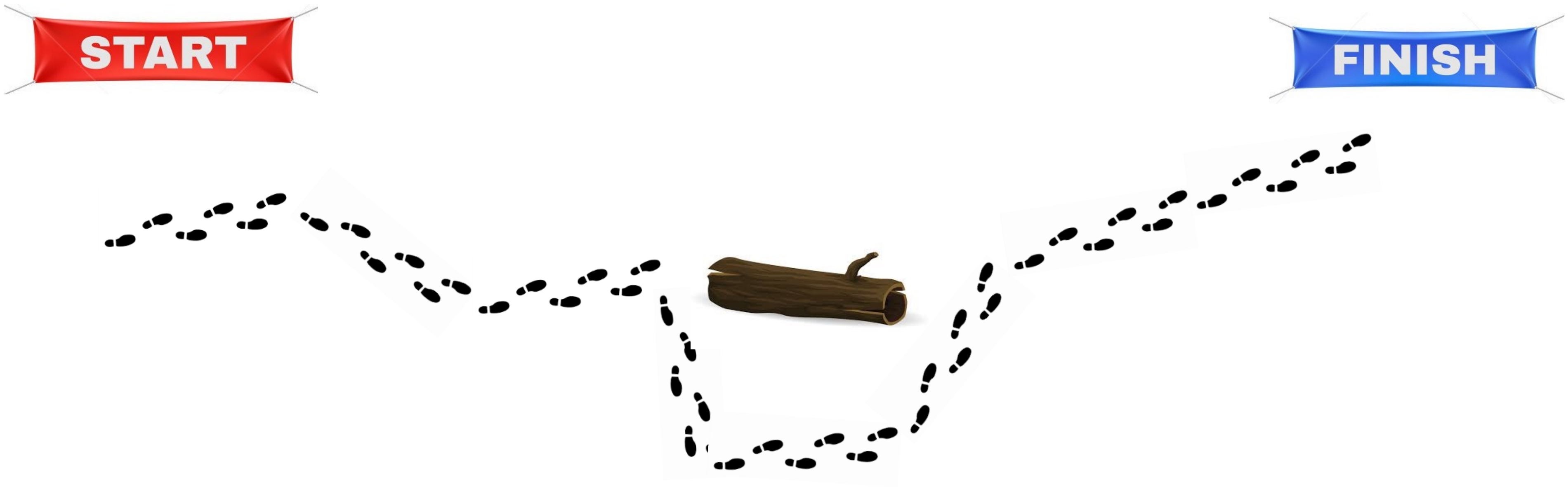


Data set

```
#> Model type: cox
#> Stratified: FALSE
#> Use covariates: FALSE
#> Status: OK
#>
#>             Estimate lower .95 upper .95     logRr seLogRr
#> treatment   0.944681  0.811174  1.096721 -0.056908  0.0769
```

Paper

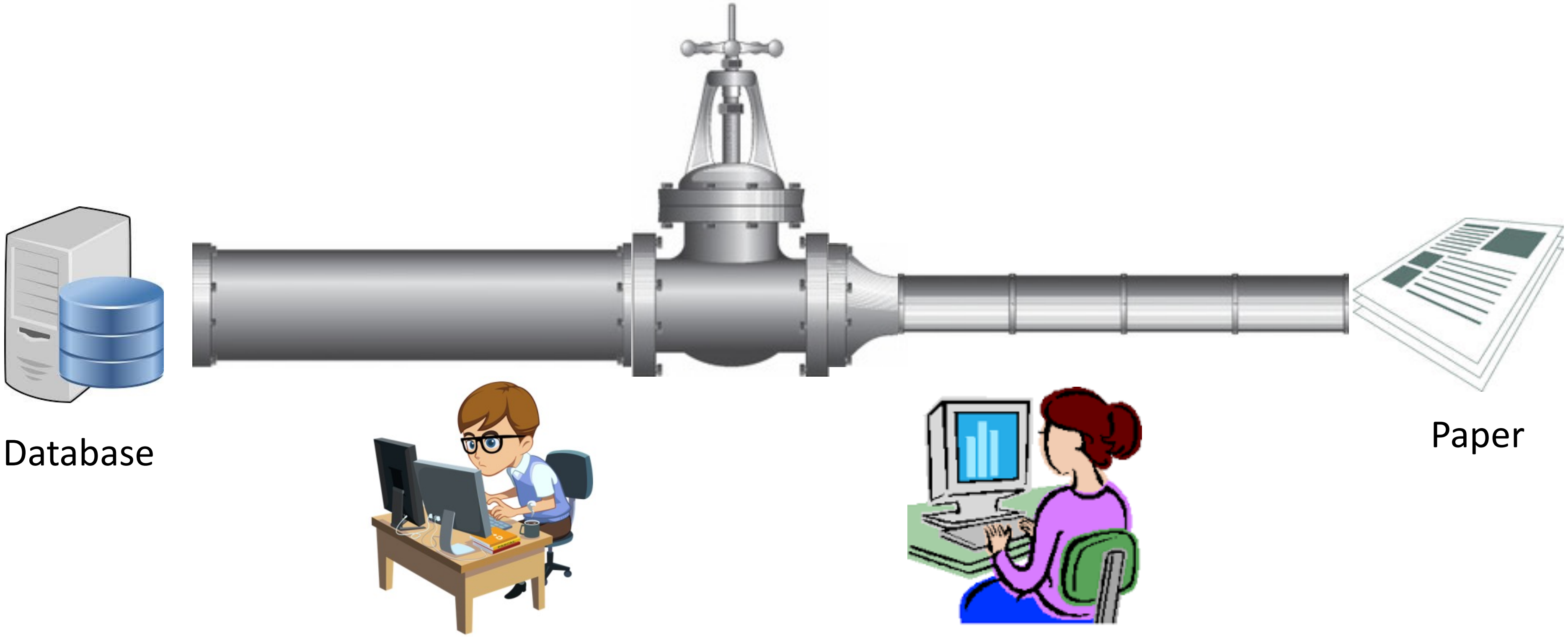
A journey from data set to paper



Most epidemiologists view a study as a journey from data set to paper.

- The protocol might be your map
- You will come across obstacles that you will have to overcome
- Several steps will require manual intervention
- In the end, it will be impossible to retrace your exact steps

What should OHDSI studies look like?



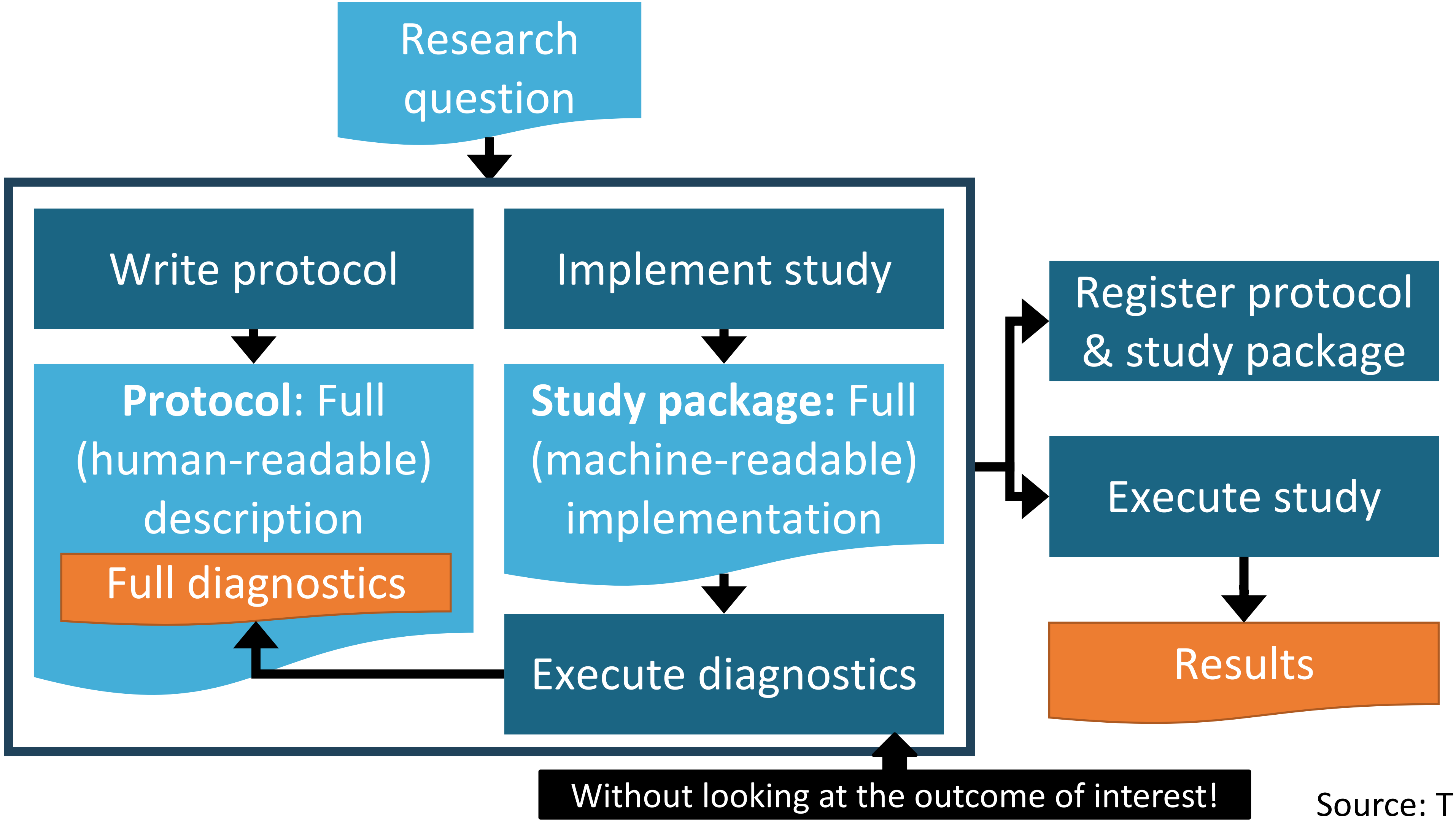
Database

Paper

A study should be like a pipeline

- A fully automated process from database to paper
- ‘Performing a study’ = building the pipeline

No peeking while developing the pipeline!

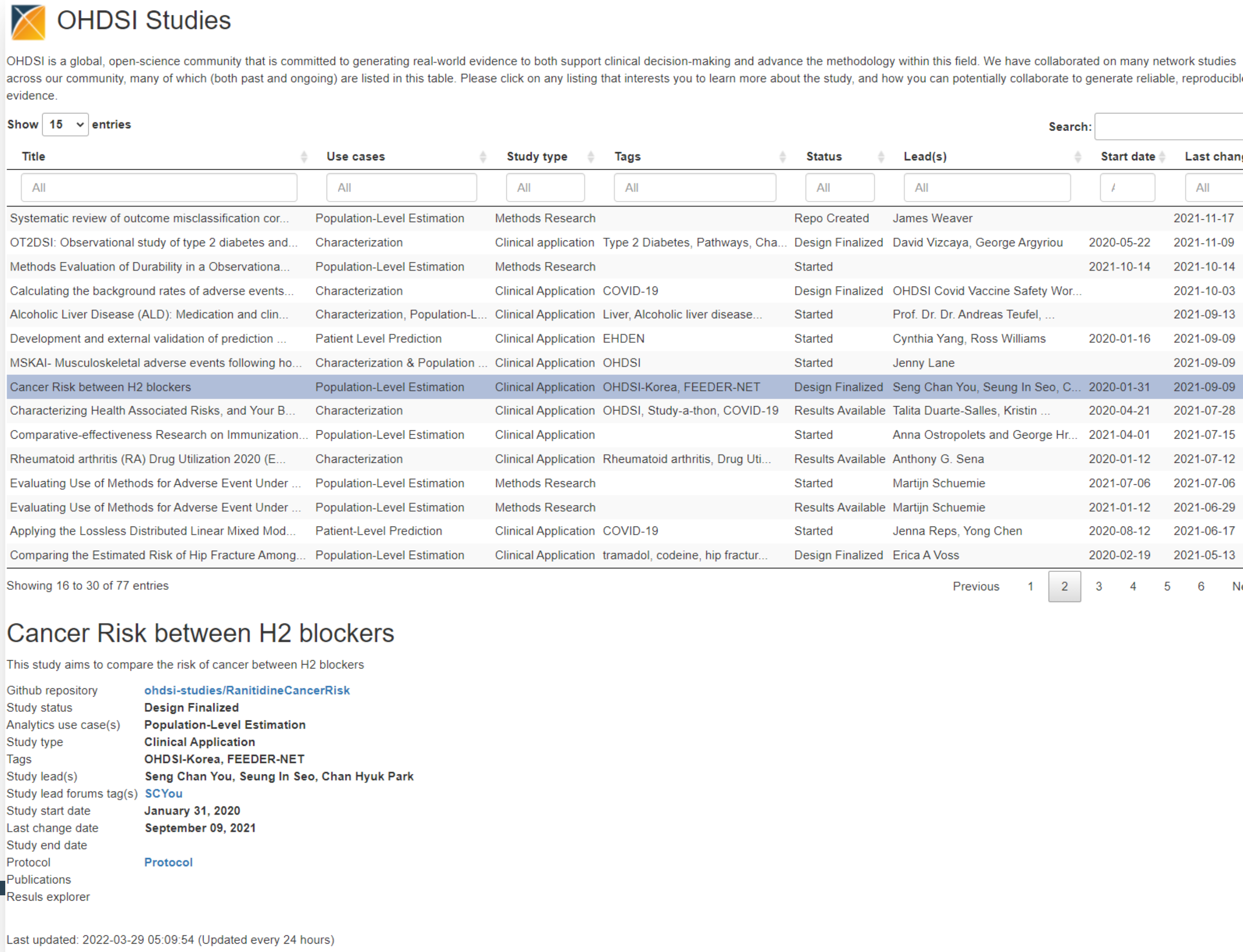

Research question
Write protocol
Implement study
Protocol: Full (human-readable) description
Full diagnostics
Study package: Full (machine-readable) implementation
Execute diagnostics
Register protocol & study package
Execute study
Results
Without looking at the outcome of interest!

Source: The Book of OHDSI

OHDSI study packages available as open source

- Each OHDSI study comes with a protocol as well as a fully executable R package
- R packages and protocols are available on

https://data.ohdsi.org/OhdsiStudies/



OHDSI Studies

OHDSI is a global, open-science community that is committed to generating real-world evidence to both support clinical decision-making and advance the methodology within this field. We have collaborated on many network studies across our community, many of which (both past and ongoing) are listed in this table. Please click on any listing that interests you to learn more about the study, and how you can potentially collaborate to generate reliable, reproducibl
evidence.

Show 15 entries

Search:

Title	Use cases	Study type	Tags	Status	Lead(s)	Start date	Last chang
All	All	All	All	All	All	/	All
Systematic review of outcome misclassification cor...	Population-Level Estimation	Methods Research		Repo Created	James Weaver		2021-11-17
OT2DSI: Observational study of type 2 diabetes and...	Characterization	Clinical application	Type 2 Diabetes, Pathways, Cha...	Design Finalized	David Vizcaya, George Argyriou	2020-05-22	2021-11-09
Methods Evaluation of Durability in a Observationa...	Population-Level Estimation	Methods Research		Started		2021-10-14	2021-10-14
Calculating the background rates of adverse events...	Characterization	Clinical Application	COVID-19	Design Finalized	OHDSI Covid Vaccine Safety Wor...		2021-10-03
Alcoholic Liver Disease (ALD): Medication and clin...	Characterization, Population-L...	Clinical Application	Liver, Alcoholic liver disease...	Started	Prof. Dr. Dr. Andreas Teufel, ...		2021-09-13
Development and external validation of prediction ...	Patient Level Prediction	Clinical Application	EHDEN	Started	Cynthia Yang, Ross Williams	2020-01-16	2021-09-09
MSKAI- Musculoskeletal adverse events following ho...	Characterization & Population ...	Clinical Application	OHDSI	Started	Jenny Lane		2021-09-09
Cancer Risk between H2 blockers	Population-Level Estimation	Clinical Application	OHDSI-Korea, FEEDER-NET	Design Finalized	Seng Chan You, Seung In Seo, C...	2020-01-31	2021-09-09
Characterizing Health Associated Risks, and Your B...	Characterization	Clinical Application	OHDSI, Study-a-thon, COVID-19	Results Available	Talita Duarte-Salles, Kristin ...	2020-04-21	2021-07-28
Comparative-effectiveness Research on Immunization...	Population-Level Estimation	Clinical Application		Started	Anna Ostropolets and George Hr...	2021-04-01	2021-07-15
Rheumatoid arthritis (RA) Drug Utilization 2020 (E...	Characterization	Clinical Application	Rheumatoid arthritis, Drug Uti...	Results Available	Anthony G. Sena	2020-01-12	2021-07-12
Evaluating Use of Methods for Adverse Event Under ...	Population-Level Estimation	Methods Research		Started	Martijn Schuemie	2021-07-06	2021-07-06
Evaluating Use of Methods for Adverse Event Under ...	Population-Level Estimation	Methods Research		Results Available	Martijn Schuemie	2021-01-12	2021-06-29
Applying the Lossless Distributed Linear Mixed Mod...	Patient-Level Prediction	Clinical Application	COVID-19	Started	Jenna Reps, Yong Chen	2020-08-12	2021-06-17
Comparing the Estimated Risk of Hip Fracture Among...	Population-Level Estimation	Clinical Application	tramadol, codeine, hip fractur...	Design Finalized	Erica A Voss	2020-02-19	2021-05-13

Showing 16 to 30 of 77 entries

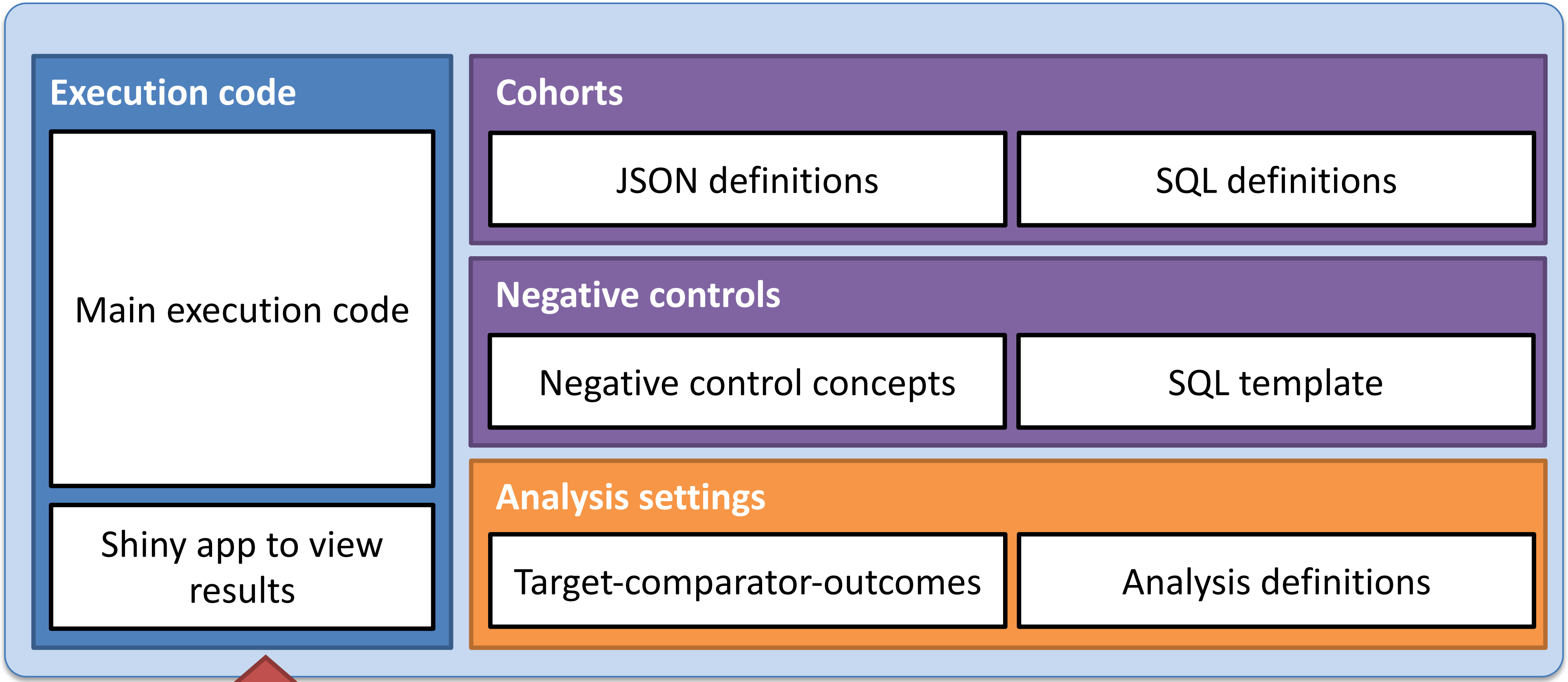
Cancer Risk between H2 blockers

This study aims to compare the risk of cancer between H2 blockers

Github repository	**ohdsi-studies/RanitidineCancerRisk**
Study status	**Design Finalized**
Analytics use case(s)	**Population-Level Estimation**
Study type	**Clinical Application**
Tags	**OHDSI-Korea, FEEDER-NET**
Study lead(s)	**Seng Chan You, Seung In Seo, Chan Hyuk Park**
Study lead forums tag(s)	**SCYou**
Study start date	**January 31, 2020**
Last change date	**September 09, 2021**
Study end date	
Protocol	**Protocol**
Publications	
Resuls explorer	

Last updated: 2022-03-29 05:09:54 (Updated every 24 hours)

Anatomy of a typical study package



Execution code

Main execution code

Shiny app to view results

Cohorts

JSON definitions

SQL definitions

Negative controls

Negative control concepts

SQL template

Analysis settings

Target-comparator-outcomes

Analysis definitions

Study code requires specific R packages to be installed. Maybe even specific R version

Preserving the compute environment

renv

Pro:

- Runs at most institutions
- Integrated in R and R-Studio
- Lightweight

Cons:

- Only preserves R package versions, not R itself
- Still being developed

docker

Pro:

- Preserves entire compute environment

Cons:

- Some OHDSI sites are not allowed to use Docker
- Docker images can get big
- Outside of R: requires additional tools + knowledge to use

Summary

- Our analyses must be 100% repeatable (on same or different data) to allow for **repeatable**, **reproducible**, **replicable**, **generalizable**, and **calibrated** science.
- Each study must be written as a **pipeline**, automatically transforming data in the **CDM** to the **study outputs** (tables, figures, etc.).
- Current OHDSI practice is to create a **study package** (in ohdsi-studies) and capture the compute environment in *renv*.
  - Exploring additional use of Docker