Reliability, replication and reproducibility: examples and perspectives

Patrick Ryan, PhD, Martijn Schuemie, PhD

Janssen Research and Development

31 Jan 2017

Reliability, reproducibility and replication

- Reliability
  - evidence can be interpreted honestly with known operating characteristics
- Reproducibility
  - Same data + same analysis = same evidence
- Replicability
  - same data + different analysis = similar evidence?
  - different data + same analysis = similar evidence?
  - different data + different analysis = similar evidence?

ORIGINAL RESEARCH | Annals of Internal Medicine

Atypical Antipsychotic Drugs and the Risk for Acute Kidney Injury and Other Adverse Outcomes in Older Adults

A Population-Based Cohort Study

Y. Joseph Hwang, MSc; Stephanie N. Dixon, PhD; Jeffrey P. Reiss, MD, MSc; Ron Wald, MD, MPH; Chirag R. Parikh, MD, PhD; Sonja Gandhi, BSc; Salimah Z. Shariff, PhD; Neesh Pannu, MD, SM; Danielle M. Nash, MSc; Faisal Rehman, MD; and Amit X. Garg, MD, PhD

Ann Intern Med. 2014;161:242-248. doi:10.7326/M13-2796

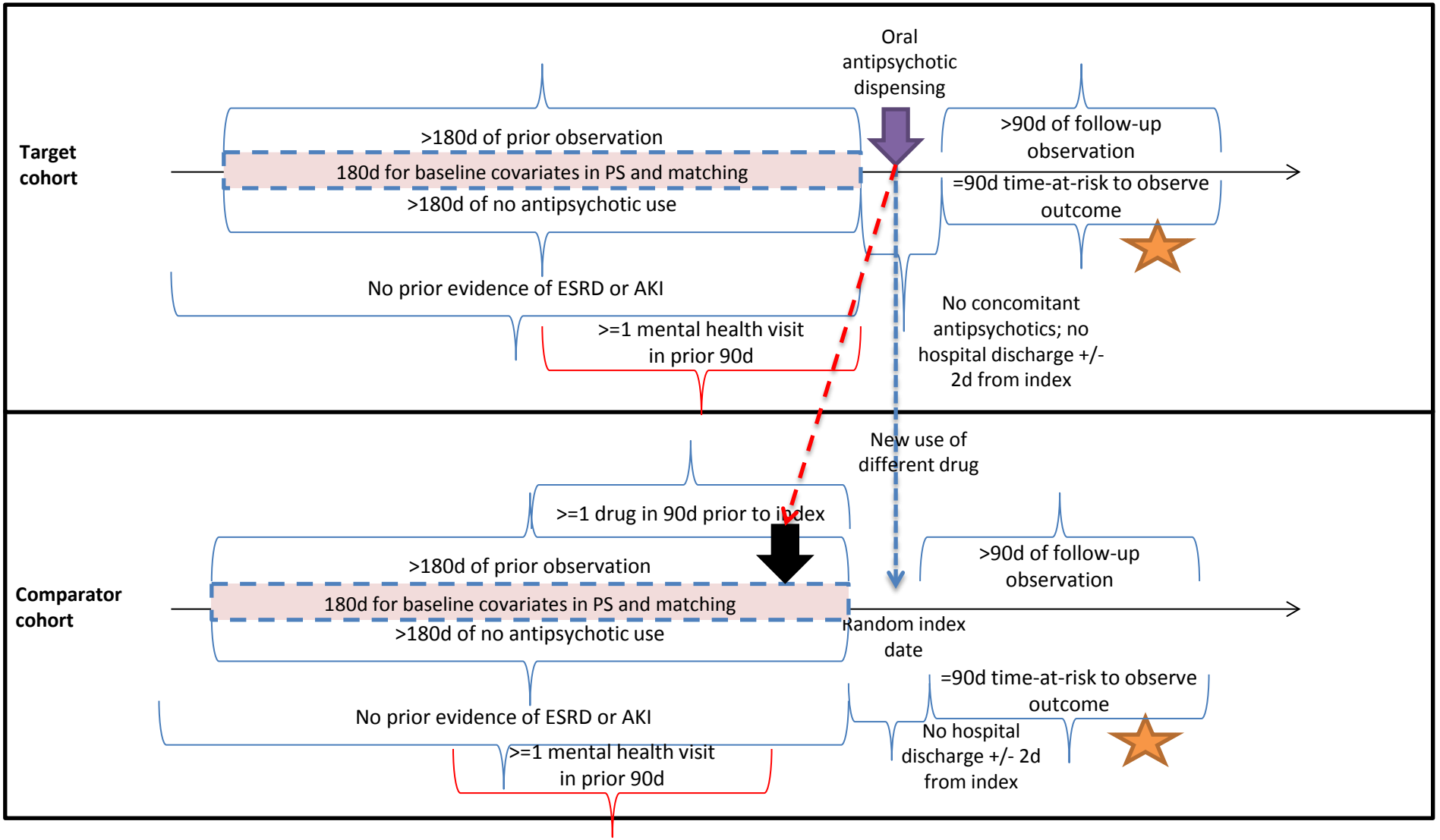
Letters

JAMA Internal Medicine Published online January 12, 2015

RESEARCH LETTER

Falls and Fractures With Atypical Antipsychotic Medication Use: A Population-Based Cohort Study

Study design issues impact performance: antipsychotic AKI/fracture story



Drugs Aging
DOI 10.1007/s40266-016-0430-x

ORIGINAL RESEARCH ARTICLE

Atypical Antipsychotics and the Risks of Acute Kidney Injury and Related Outcomes Among Older Adults: A Replication Analysis and an Evaluation of Adapted Confounding Control Strategies

Patrick B. Ryan[1] · Martijn J. Schuemie[1] · Darmendra Ramcharran[1] · Paul E. Stang[1]

ORIGINAL CONTRIBUTION

Journal of Clinical Psychopharmacology • Volume 37, Number 2, April 2017

Atypical Antipsychotics and the Risk of Falls and Fractures Among Older Adults

An Emulation Analysis and an Evaluation of Additional Confounding Control Strategies

Darmendra Ramcharran, PhD, Hong Qiu, MD, PhD, Martijn J. Schuemie, PhD, and Patrick B. Ryan, PhD

Table 3 Replication of the Hwang et al. [1] model and adapted analyses (exposure group: new user of any atypical antipsychotic)

90-Day hospitalization event	Model	Exposure events, *n* (%)	Comparator events, *n* (%)	OR	95% CI	Theoretical *p* value	Empirical *p* value
Acute kidney injury	Hwang effect estimate	1002 (1.02)	602 (0.62)	1.73	1.55–1.92	NS	NS
	Replication	1043 (1.07)	717 (0.74)	1.45	1.32–1.60	<0.01	0.41
	Adapted[a]	373 (1.13)	420 (1.27)	0.91	0.78–1.07	0.26	0.91
Hypotension	Hwang effect estimate	384 (0.39)	215 (0.22)	1.91	1.60–2.28	NS	NS
	Replication	686 (0.73)	420 (0.45)	1.63	1.45–1.85	<0.01	0.22
	Adapted[a]	253 (0.8)	263 (0.83)	1.03	0.86–1.24	0.74	0.23
Acute urinary retention	Hwang effect estimate	329 (0.34)	170 (0.17)	1.98	1.63–2.40	NS	NS
	Replication	322 (0.34)	197 (0.21)	1.63	1.37–1.95	<0.01	0.23
	Adapted[a]	124 (0.38)	119 (0.37)	1.09	0.84–1.41	0.53	0.20
Neuroleptic malignant syndrome or rhabdomyolysis	Hwang effect estimate	99 (0.10)	69 (0.07)	1.36	0.96–1.62	NS	NS
	Replication	89 (0.09)	33 (0.03)	2.70	1.83–4.08	<0.01	0.01
	Adapted[a]	31 (0.09)	26 (0.08)	1.19	0.71–2.02	0.51	0.32
Pneumonia							NS
							0.31
Acute myocardi							0.28
							NS
							0.93
Ventricular arrh							0.08
							NS
							0.81
	Adapted[a]	62 (0.19)	69 (0.21)	0.93	0.63–1.37	0.72	0.88
Death (in-hospital)	Hwang effect estimate	6666 (6.82)	2985 (3.05)	2.39	2.28–2.50	NS	NS
	Replication	273 (0.28)	145 (0.15)	1.88	1.54–2.31	<0.01	0.10
	Adapted[a]	60 (0.18)	157 (0.47)	0.38	0.28–0.51	<0.01	<0.01

CI confidence interval, *NS* not specified, *OR* odds ratio

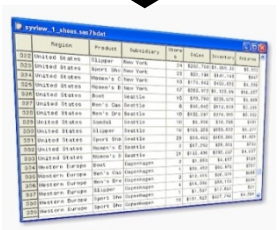
[a] The final logistic regression fit by Hwang et al. [1] with a requirement for patients to have a diagnosis of schizophrenia, bipolar disorder, or major depression, a healthcare visit within 90 days prior to the index date, and additional adjustment for all covariates entered into the propensity score model

Lessons:

- Challenge in reproducibility
- Value in replication: same analysis, different data
- Value in negative controls to assess reliability
- Value in different analyses to evaluate robustness

Study reproducibility

What do epi studies currently look like?

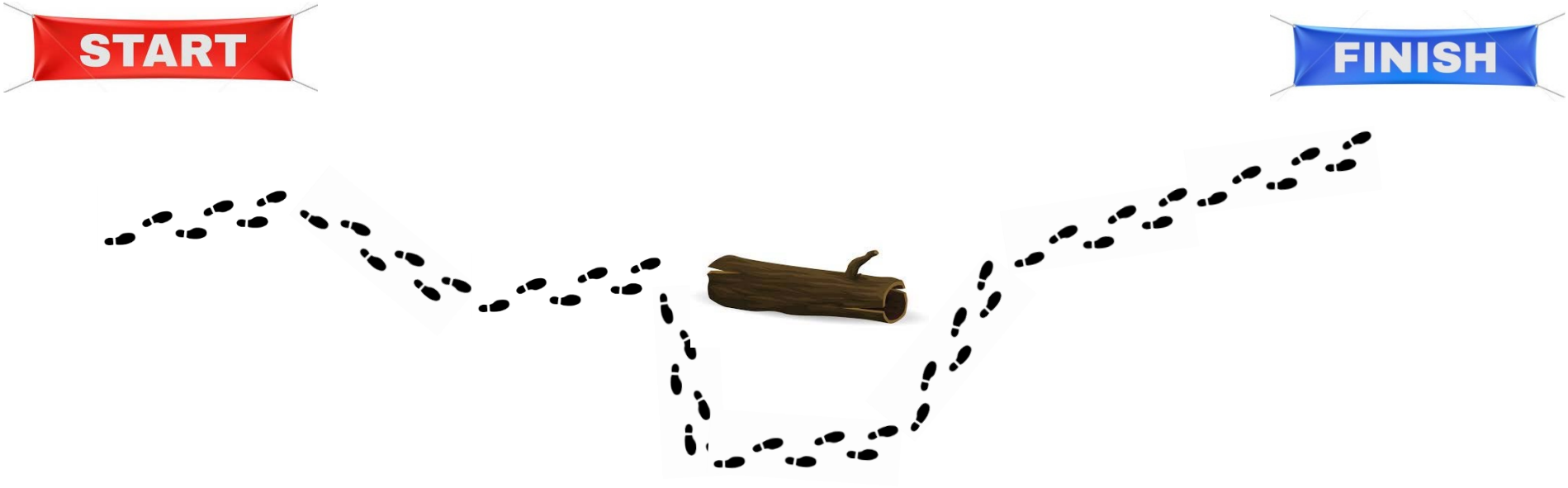


Data set

```
#> 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



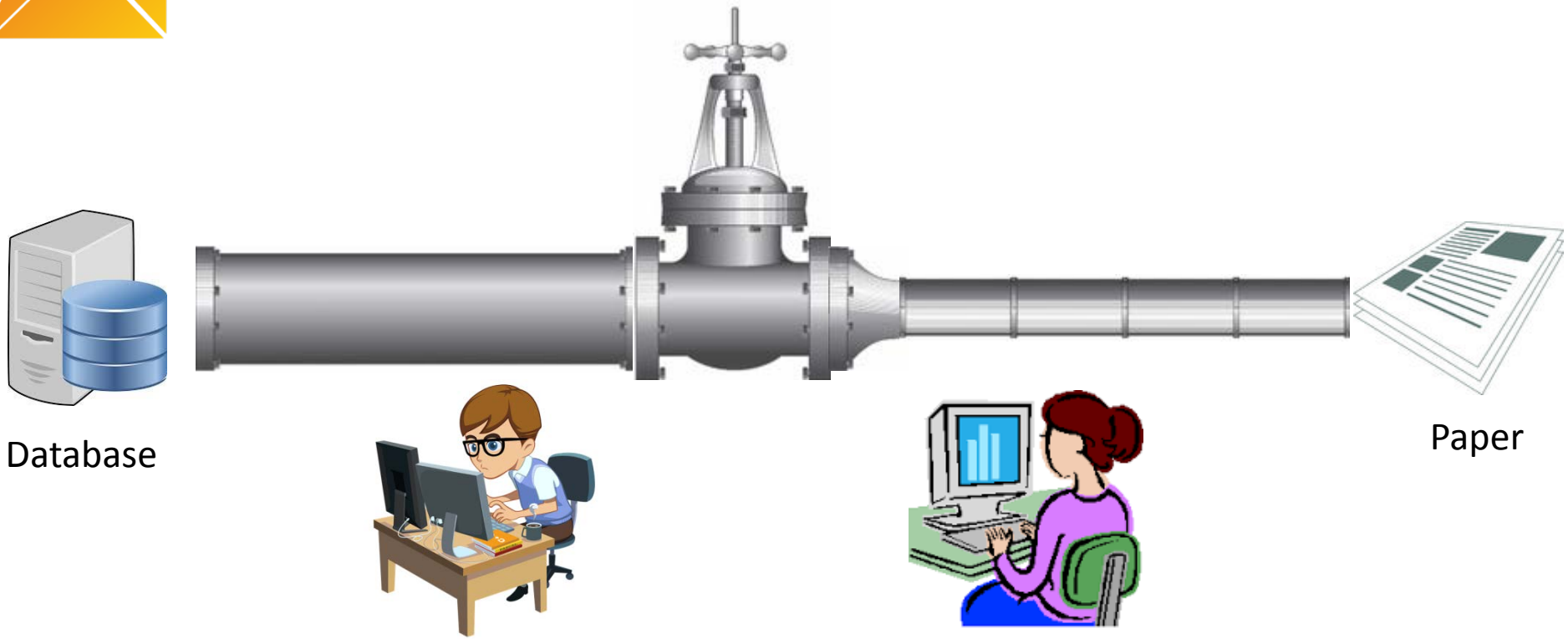
Most epidemiologists view a study as a journey from data set to paper.

- 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

Current epi studies are non-reproducible

- How do we know what happened?
- How do we know if it was done correctly?
- How do we know how well it worked?
- How could we be more efficient?
- How can we deal with more complex studies?
- How can multiple people work together on the same analysis?
- How could other reproduce this study on a different database?

What should OHDSI studies look like?



A study should be like a pipeline

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

Example: Keppra – angioedema study

OHDSI study:

- Does exposure to Keppra (levetiracetam) lead to an increased risk of angioedema?
- Compared to phenytoin

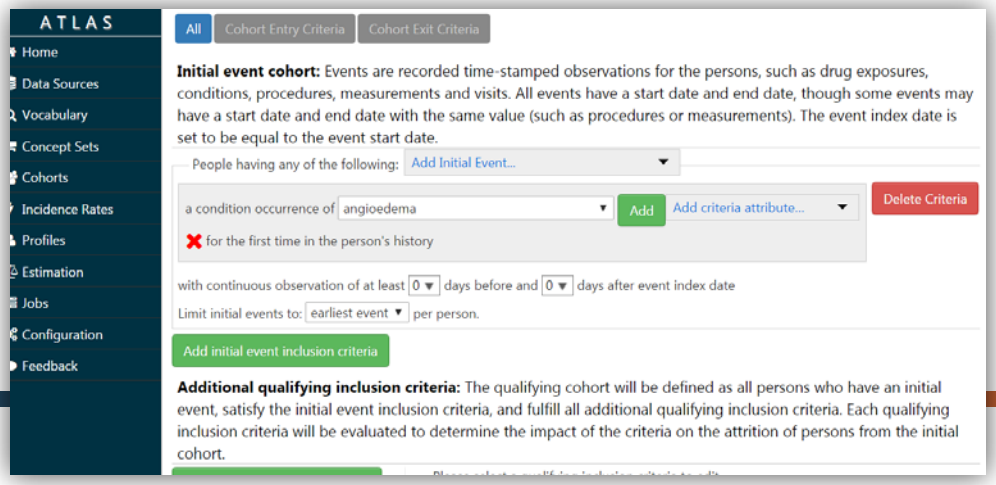
https://github.com/OHDSI/StudyProtocols/tree/master/KeppraAngioedema

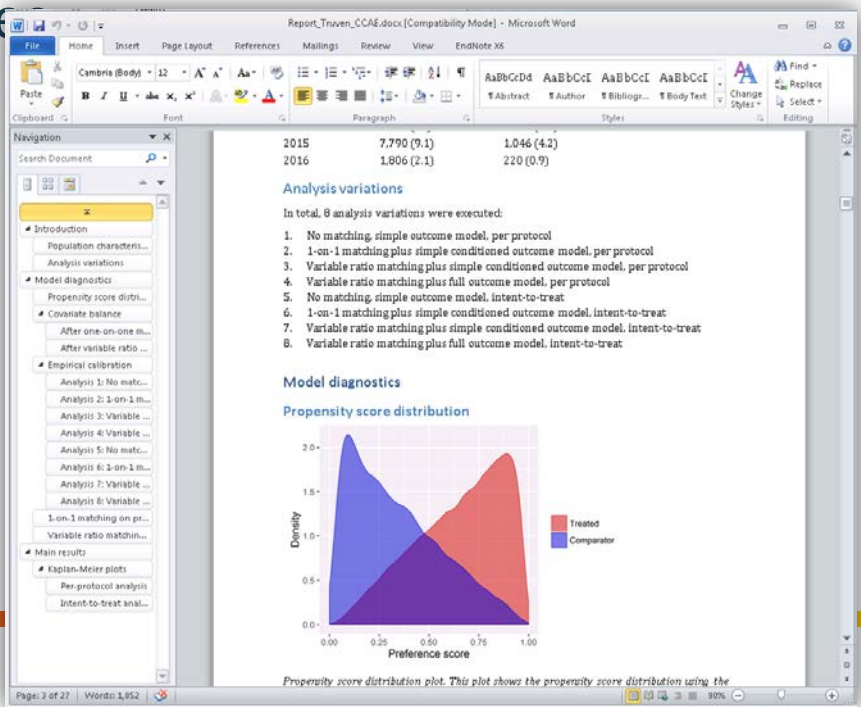
```
library(KeppraAngioedema)

connectionDetails <- createConnectionDetails(dbms = "postgresql",
                                             user = "joe",
                                             password = "secret",
                                             server = "myserver")

execute(connectionDetails,
        cdmDatabaseSchema = "cdm_data",
        workDatabaseSchema = "results",
        studyCohortTable = "ohdsi_keppra_angioedema",
        oracleTempSchema = NULL,
        outputFolder = "c:/temp/study_results",
        maxCores = 4)
```

Full traceability

Study package contains

- Cohort definitions (e.g. angioedema definition)
  - OhdsiRTools::insertCohortDefinitionInPackage(2193, "Angioedema")
- All analysis details for the CohortMethod package
- CohortMethod package describes data extraction
- Code to generate tables and figure
- Code to generate full report





Full traceability

R environment snapshot

```
OhdsiRTools::insertEnvironmentSnapshotInPackage("KeppraAngioedema")
```

81 lines (80 sloc) | 1.42 KB — Raw | Blame | History

Search this file...

	package	version
2	R	3.3.1
3	grDevices	3.3.1
4	graphics	3.3.1
5	utils	3.3.1
6	stats	3.3.1
7	bit	1.1-12
8	methods	3.3.1
9	tools	3.3.1
10	colorspace	1.2-6
11	DBI	0.5
12	fastmatch	1.0-4
13	ff	2.2-13
14	grid	3.3.1
15	magrittr	1.5
16	Rcpp	0.12.8
17	rJava	0.9-8
18	stringi	1.1.2
19	assertthat	0.1
20	curl	1.2

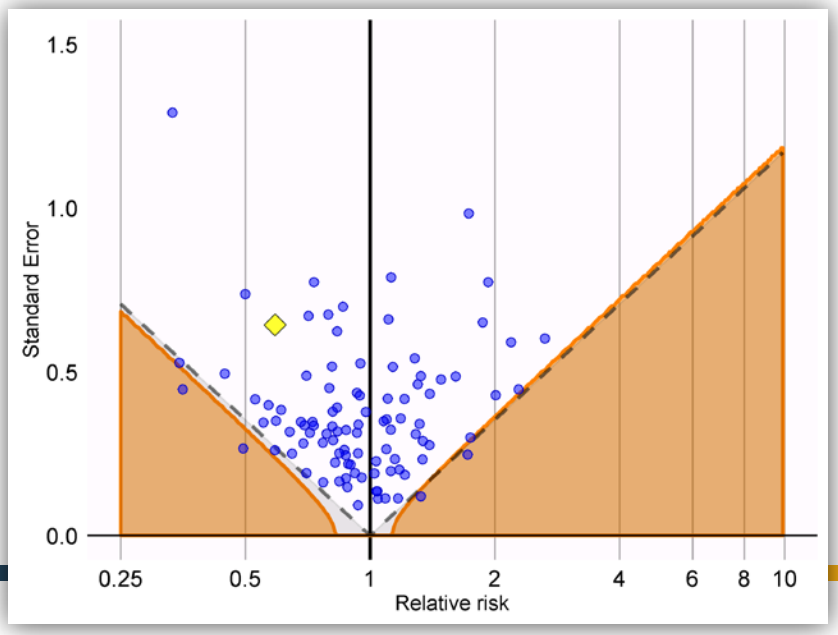
We can check for correctness

- We can review the study code
- We should make the study code publicly available as part of the paper
- Large parts of the study are automatically checked using unit tests

```
test_that("Simple 1-on-1 matching", {
  rowId <- 1:5
  treatment <- c(1, 0, 1, 0, 1)
  propensityScore <- c(0, 0.1, 0.3, 0.4, 1)
  data <- data.frame(rowId = rowId, treatment = treatment, propensityScore = propensityScore)
  result <- matchOnPs(data, caliper = 0, maxRatio = 1)
  expect_equal(result$stratumId, c(0, 0, 1, 1))
})

test_that("Simple 1-on-n matching", {
  rowId <- 1:6
  treatment <- c(0, 1, 0, 0, 1, 0)
  propensityScore <- c(0, 0.1, 0.12, 0.85, 0.9, 1)
  data <- data.frame(rowId = rowId, treatment = treatment, propensityScore = propensityScore)
  result <- matchOnPs(data, caliper = 0, maxRatio = 100)
  expect_equal(result$stratumId, c(0, 0, 0, 1, 1, 1))
})
```

We can evaluate how well the study worked

- Included 100 negative control outcomes
- Results show little residual confounding when using propensity score matching



Writing the study was very efficient

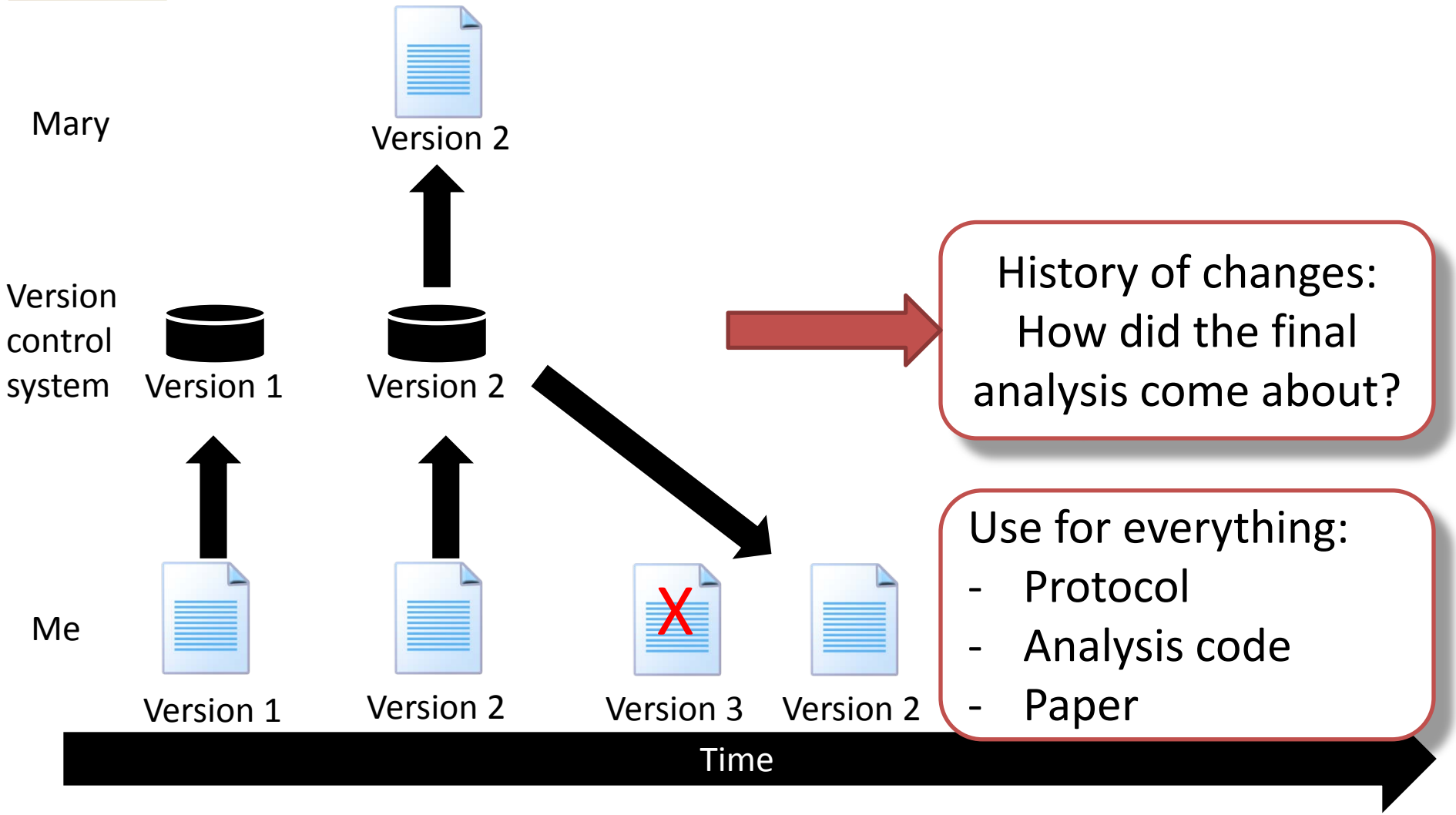
- Reuse of R code in CohortMethod, DatabaseConnector, SqlRender, EmpiricalCalibration, etc.
- Implementation took days instead of months
- Next study will be faster

Complexity is not a problem

Use software engineering approaches to deal with complexity:

- Abstraction
- Encapsulation
- Writing clear code
- Re-use

Several people can work on the same analysis through version control



Commit log

6 PGxDrugStudy/inst/sql/sql_server/CountGender.sql

View

```
@@ -1,4 +1,4 @@
-# query-to-get-count-of-males-and-females-being-prescribed-any-drug-using-age-at-exposure
+-- query-to-get-count-of-males-and-females-being-prescribed-any-drug-using-age-at-exposure

 SELECT CONCEPT.concept_name as gender, COUNT(DISTINCT(PERSON.person_id))
 FROM DRUG_EXPOSURE,  PERSON, CONCEPT
@@ -6,8 +6,8 @@ WHERE DRUG_EXPOSURE.DRUG_EXPOSURE_START_DATE >= DATE '2009-01-01'
  AND   DRUG_EXPOSURE.DRUG_EXPOSURE_START_DATE <= DATE '2012-12-31'
  AND   DRUG_EXPOSURE.person_id = PERSON.person_id
  AND   PERSON.gender_concept_id = CONCEPT.concept_id
-AND   (DATE_PART_YEAR(DRUG_EXPOSURE.DRUG_EXPOSURE_START_DATE) - PERSON.year_of_birth >= 0)
-AND   (DATE_PART_YEAR(DRUG_EXPOSURE.DRUG_EXPOSURE_START_DATE) - PERSON.year_of_birth < 14)
+AND   (YEAR(DRUG_EXPOSURE.DRUG_EXPOSURE_START_DATE) - PERSON.year_of_birth >= 0)
+AND   (YEAR(DRUG_EXPOSURE.DRUG_EXPOSURE_START_DATE) - PERSON.year_of_birth < 14)
 GROUP BY gender
 ORDER BY gender
```

Easy to rerun on different data

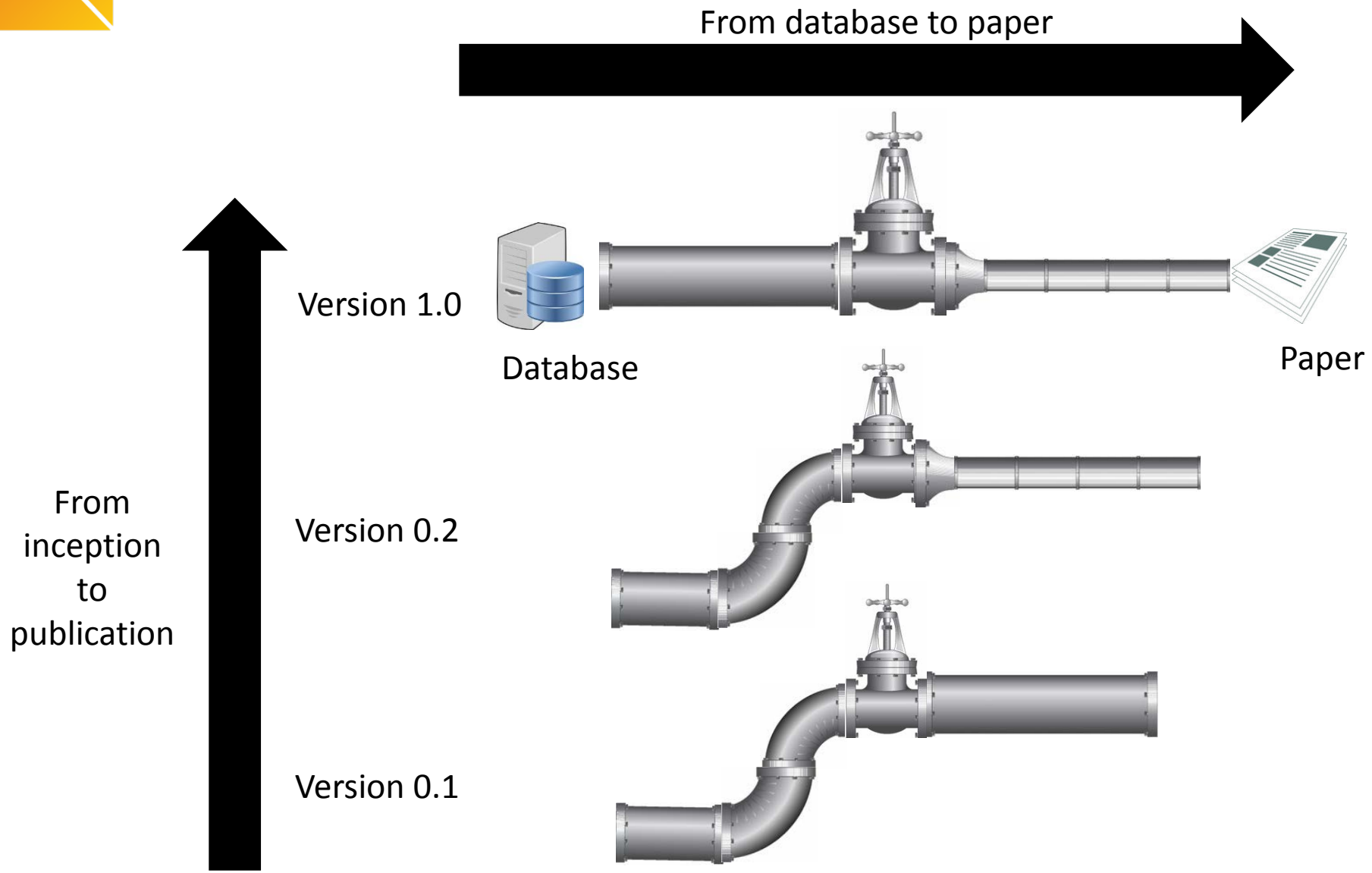
The Keppra – Angioedema study was run on:

- Columbia University EHR
- Stanford EHR
- Cerner (University of Texas)
- Pharmetrics Plus (IMS)
- Optum
- Truven CCAE
- Truven MDCD
- Truven MDCR
- …

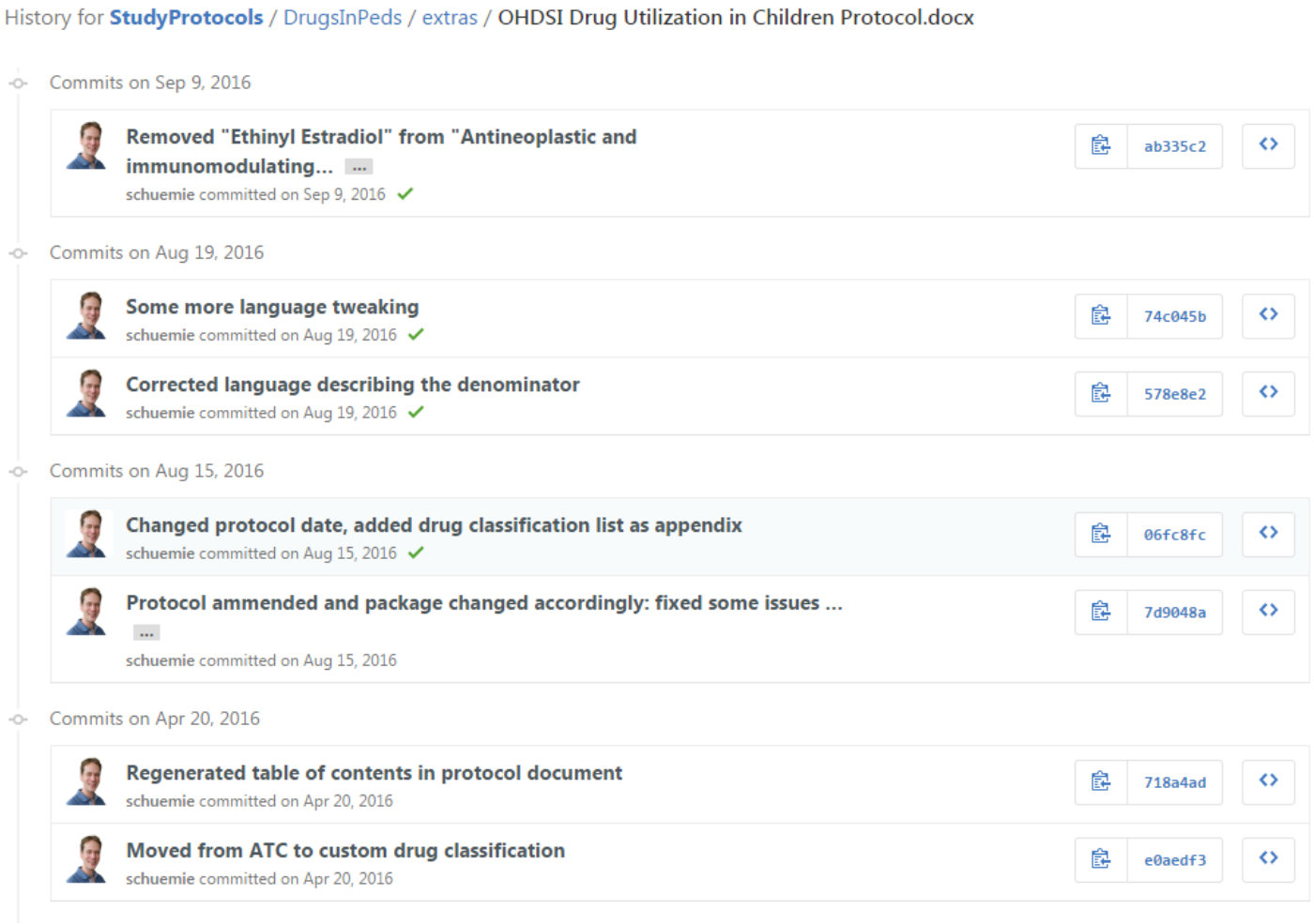
Viewing a study as a pipeline has many advantages

- Full traceability
- Ability to check for correctness
- Ability to evaluate using controls
- More efficient
- Ability to deal with complexity
- Ability to work with several people on one analysis
- Easy to rerun on different data

Two dimensions of reproducibility



Version control supports the 2nd dimension



Conclusions

- Most epi studies lack reproducibility
- 1st dimension: From database to paper
- 2nd dimension: From inception to publication
- Studies should be viewed as pipelines
- The pipeline should be published as part of the paper

Join the journey

- Discussion / questions / comments

ryan@ohdsi.org