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?
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


Letters

RESEARCH LETTER

Falls and Fractures With Atypical Antipsychotic Medication Use: A Population-Based Cohort Study
Study design issues impact performance: antipsychotic AKI/fracture story

Target cohort:
- >180d of prior observation
- 180d for baseline covariates in PS and matching
- >180d of no antipsychotic use
- No prior evidence of ESRD or AKI
- >=1 mental health visit in prior 90d

Comparator cohort:
- >180d of prior observation
- 180d for baseline covariates in PS and matching
- >180d of no antipsychotic use
- No prior evidence of ESRD or AKI
- >=1 mental health visit in prior 90d

Oral antipsychotic dispensing:
- >90d of follow-up observation
- =90d time-at-risk to observe outcome
- No concomitant antipsychotics; no hospital discharge +/- 2d from index

Random index date:
- >=1 drug in 90d prior to index
- New use of different drug

Comparative analysis:
- No hospital discharge +/- 2d from index
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. Ryan¹ · Martijn J. Schuemie¹ · Darmendra Ramcharran¹ · Paul E. Stang¹

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
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

<table>
<thead>
<tr>
<th>90-Day hospitalization event</th>
<th>Model</th>
<th>Exposure events, n (%)</th>
<th>Comparator events, n (%)</th>
<th>OR</th>
<th>95% CI</th>
<th>Theoretical p value</th>
<th>Empirical p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute kidney injury</td>
<td>Hwang effect estimate</td>
<td>1002 (1.02)</td>
<td>602 (0.62)</td>
<td>1.73</td>
<td>1.55–1.92</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>Replication</td>
<td>1043 (1.07)</td>
<td>717 (0.74)</td>
<td>1.45</td>
<td>1.32–1.60</td>
<td>&lt;0.01</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>Adapted</td>
<td>373 (1.13)</td>
<td>420 (1.27)</td>
<td>0.91</td>
<td>0.78–1.07</td>
<td>0.26</td>
<td>0.91</td>
</tr>
<tr>
<td>Hypotension</td>
<td>Hwang effect estimate</td>
<td>384 (0.39)</td>
<td>215 (0.22)</td>
<td>1.91</td>
<td>1.60–2.28</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>Replication</td>
<td>686 (0.73)</td>
<td>420 (0.45)</td>
<td>1.63</td>
<td>1.45–1.85</td>
<td>&lt;0.01</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>Adapted</td>
<td>253 (0.8)</td>
<td>263 (0.83)</td>
<td>1.03</td>
<td>0.86–1.24</td>
<td>0.74</td>
<td>0.23</td>
</tr>
<tr>
<td>Acute urinary retention</td>
<td>Hwang effect estimate</td>
<td>329 (0.34)</td>
<td>170 (0.17)</td>
<td>1.98</td>
<td>1.63–2.40</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>Replication</td>
<td>322 (0.34)</td>
<td>197 (0.21)</td>
<td>1.63</td>
<td>1.37–1.95</td>
<td>&lt;0.01</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>Adapted</td>
<td>124 (0.38)</td>
<td>119 (0.37)</td>
<td>1.09</td>
<td>0.84–1.41</td>
<td>0.53</td>
<td>0.20</td>
</tr>
<tr>
<td>Neuroleptic malignant syndrome or rhabdomyolysis</td>
<td>Hwang effect estimate</td>
<td>99 (0.10)</td>
<td>69 (0.07)</td>
<td>1.36</td>
<td>0.96–1.62</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>Replication</td>
<td>89 (0.09)</td>
<td>33 (0.03)</td>
<td>2.70</td>
<td>1.83–4.08</td>
<td>&lt;0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Adapted</td>
<td>31 (0.09)</td>
<td>26 (0.08)</td>
<td>1.19</td>
<td>0.71–2.02</td>
<td>0.51</td>
<td>0.32</td>
</tr>
</tbody>
</table>

CI confidence interval, NS not specified, OR odds ratio

* 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.
Study reproducibility
What do epi studies currently look like?

Data set → SPSS → Excel → 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
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 figures
- Code to generate full report
Full traceability

R environment snapshot

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

```r
# 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, 0.05, 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.

Version control system

Mary

Version 1

Version 2

Version 2

Version 3

Version 1

Version 2

History of changes:
How did the final analysis come about?

Use for everything:
- Protocol
- Analysis code
- Paper
Commit log

```sql
SELECT CONCEPT.concept_name as gender, COUNT(DISTINCT(PERSON.person_id))
FROM drug_exposure, person, concept
WHERE drug_exposure.drug_exposure_start_date >= DATE '2009-01-01'
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

From database to paper

Version 1.0
Database

Version 0.2

Version 0.1

From inception to publication
Version control supports the 2\textsuperscript{nd} dimension

<table>
<thead>
<tr>
<th>History for StudyProtocols / DrugsInPeds / extras / OHDSI Drug Utilization in Children Protocol.docx</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Commits on Sep 9, 2016</strong></td>
</tr>
<tr>
<td>- Removed &quot;Ethinyl Estradiol&quot; from &quot;Antineoplastic and immunomodulating...&quot;</td>
</tr>
<tr>
<td>- schuemle committed on Sep 9, 2016</td>
</tr>
<tr>
<td><strong>Commits on Aug 19, 2016</strong></td>
</tr>
<tr>
<td>- Some more language tweaking</td>
</tr>
<tr>
<td>- schuemle committed on Aug 19, 2016</td>
</tr>
<tr>
<td>- Corrected language describing the denominator</td>
</tr>
<tr>
<td>- schuemle committed on Aug 19, 2016</td>
</tr>
<tr>
<td><strong>Commits on Aug 15, 2016</strong></td>
</tr>
<tr>
<td>- Changed protocol date, added drug classification list as appendix</td>
</tr>
<tr>
<td>- schuemle committed on Aug 15, 2016</td>
</tr>
<tr>
<td>- Protocol amended and package changed accordingly: fixed some issues ...</td>
</tr>
<tr>
<td>- schuemle committed on Aug 15, 2016</td>
</tr>
<tr>
<td><strong>Commits on Apr 20, 2016</strong></td>
</tr>
<tr>
<td>- Regenerated table of contents in protocol document</td>
</tr>
<tr>
<td>- schuemle committed on Apr 20, 2016</td>
</tr>
<tr>
<td>- Moved from ATC to custom drug classification</td>
</tr>
<tr>
<td>- schuemle committed on Apr 20, 2016</td>
</tr>
</tbody>
</table>
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