Using Healthcare Databases To Evaluate The Safety And Effectiveness Of Newly Marketed Medications

Sebastian Schneeweiss, MD, ScD
Professor of Medicine and Epidemiology

Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine
Brigham and Women’s Hospital, Harvard Medical School, Boston

March 2018
Funding

- This study was funded in part by FDA HHSF223201710186C
- This study was funded in part by the Division of Pharmacoepidemiology
- In addition, Dr. Schneeweiss was funded by PCORI

Disclosures

- PI, Harvard-Brigham & Women’s Hospital Drug Safety Research Center (FDA)
- Co-Chair, Methods Core of the FDA Sentinel System
- PI of research grants awarded to BWH by Bayer, Genentech, Boehringer Ingelheim
- Consulting fees from WHISCON, LLC, and Aetion, Inc. (incl. equity)
- Grants/contracts from NIH, AHRQ, PCORI, FDA, IMI, Arnold Foundation
21st Century Cures Act and PDUFA VI: The role of RWE

FDA debates the utility of Real-World Evidence

RWD: Routine data from healthcare systems

Real-World Evidence — What Is It and What Can It Tell Us?

Rachel E. Sherman, M.D., M.P.H., Steven A. Anderson, Ph.D., M.P.P.,
Gerald J. Dal Pan, M.D., M.H.S., Gerry W. Gray, Ph.D., Thomas Gross, M.D., M.P.H.,
Nina L. Hunter, Ph.D., Lisa LaVange, Ph.D., Danica Marinac-Dabic, M.D., Ph.D.,
Peter W. Marks, M.D., Ph.D., Melissa A. Robb, B.S.N., M.S., Jeffrey Shuren, M.D., J.D.,
Robert Temple, M.D., Janet Woodcock, M.D., Lilly Q. Yue, Ph.D., and Robert M. Califf, M.D.

Multidimensional Evidence Generation and FDA Regulatory Decision Making
Defining and Using “Real-World” Data

Jonathan P. Jarow, MD  Lisa LaVange, PhD  Janet Woodcock, MD
Effectiveness Research with Healthcare Databases

RCT data

Non-interventional data

10%

Research data

Data collected PRIMARILY for research

For purpose

Data specifically for study purpose

Other purpose

Data intended for other studies

90%

Transactional data

Data used SECONDARILY for research

Other purpose

Clinical documentation

Administrative

Database Studies

- Framingham Study
- Cardiovas Health Study
- Slone Birth Defects Study
- Some registries

- Nurses’ Health Study 1
- Some registries

- EHR-based studies
- NDI linkage
- Lab test databases
- Some registries

- Claims data studies
- Geocoding/census

Example

- Some registries
From transactional data to study implementation

A. Dynamic database that records an ongoing stream of new healthcare records in Calendar Time for all enrolled patients.

B. Stabilized data snapshot for research purposes.

C. Individual-patient data has arrived in episodes and from various sources.

D. Study rules are applied and arranged by Event Time.

Healthcare records are entered as they arrive, sorted by service date. (Some records arrive with admin delays.)
RWE in regulatory decision making: Key use cases

1. Secondary indications
   - Approval
   - 2nd Indication
   - Exps: Pediatric, other endpoints, other disease stages

2. Adaptive Pathways
   - Initial Approval
   - Full Approval
   - Exps: Biomarker to clinical endpoint, broader popn

3. Safety (a)
   - Approval
   - Exps: Post-market requirements (PMR), rapid regulatory response

Safety (b)
A Comparison of Aprotinin and Lysine Analogues in High-Risk Cardiac Surgery

Risk of death (7d)
HR = 1.78 (1.56 -2.02)

Risk of death (30d)
HR = 1.53 (1.06 -2.22)
**Database Study**

**ARTHRITIS & RHEUMATOLOGY**

Cardiovascular Safety of Tocilizumab Versus Tumor Necrosis Factor Inhibitors in Patients With Rheumatoid Arthritis

A Multi-Database Cohort Study

Seoung C. Kim,1 Daniel H. Solomon,1 James R. Rogers,1 Sara Gale,2 Micki Klearman,2 Khaled Sarsour,2 and Sebastian Schneeweiss1

**Risk of composite CV outcome**

HR = 0.85 (0.61-1.19)

---

**RCT**

**ABSTRACT NUMBER: 3L**

Comparative Cardiovascular Safety of Tocilizumab Vs Etanercept in Rheumatoid Arthritis: Results of a Randomized, Parallel-Group, Multicenter, Noninferiority, Phase 4 Clinical Trial

Jon T. Giles1, Naveed Sattar4, Sherine E. Gabriel1, Paul M. Ridker5, Steffen Gay2, Charles Warner6, David Musselman7, Laura Brockwell6, Emma Shittu6, Micki Klearman2 and Thomas Fleming8

**ENTRACTE**

**Risk of composite CV outcome**

HR = 1.05 (0.77-1.43)

---

**TCZ**

<table>
<thead>
<tr>
<th>As-treated analysis</th>
<th>No. of subjects</th>
<th>No. of events</th>
<th>Person-years</th>
<th>IR (95% CI)†</th>
<th>HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite cardiovascular events Medicare</td>
<td>2,531</td>
<td>17</td>
<td>1,841</td>
<td>0.92 (0.56-1.44)</td>
<td>0.70 (0.40-1.24)</td>
</tr>
<tr>
<td>PharMetrics</td>
<td>2,614</td>
<td>10</td>
<td>2,061</td>
<td>0.49 (0.25-0.86)</td>
<td>1.00 (0.45-2.22)</td>
</tr>
<tr>
<td>MarketScan</td>
<td>4,073</td>
<td>9</td>
<td>2,999</td>
<td>0.30 (0.15-0.55)</td>
<td>1.03 (0.46-2.34)</td>
</tr>
<tr>
<td>Combined</td>
<td>9,218</td>
<td>36</td>
<td>6,901</td>
<td>0.52 (0.37-0.71)</td>
<td>0.84 (0.56-1.26)</td>
</tr>
</tbody>
</table>

**First Events, n**

<table>
<thead>
<tr>
<th>Etanercept N = 1542</th>
<th>Tocilizumab N = 1538 vs Etanercept</th>
<th>Tocilizumab N = 1538 vs Etanercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Events, n</td>
<td>78</td>
<td>83</td>
</tr>
</tbody>
</table>

© 2018 Harvard / Brigham Division of Pharmacoepidemiology
Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes

**RCT Database Study**

Followed by **Database Study**

**Empagliflozin and risk of DKA**

1 / 2,333 vs. 3 / 2,345

HR = 2.9 (0.4-20.0)

**SGLT-2 and risk of DKA**

26 / 38,045 vs. 55 / 38,045

HR = 2.2 (1.4-3.6)
RCT followed by Database Study

**RE-LY**

Stroke prevention
HR = 0.66 (0.53-0.82)

Stroke prevention
HR = 0.77 (0.54-1.09)
Prevention of heart failure hospitalization
HR = 0.61 (0.47-0.78)

Prevention of heart failure hospitalization
HR = 0.67 (0.52-0.87)

Database Study followed by RCT

Original Article

Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes

Bruce Neal, M.B., Ch.B., Ph.D., Viola Ferkovic, M.B., B.S., Ph.D., Kenneth W. Mahaffey, M.D., Dick de Zeeuw, M.D., Ph.D., Craig Fulcher, M.D., Ngozi Enodu, M.D., Mehul Desai, M.D., Ph.D., Jun Liu, Seoyoung C Kim

Cardiovascular outcomes associated with canagliflozin versus other non-gliflozin antidiabetic drugs: population based cohort study

Elisabetta Patomo, Allison B Goldfine, Sebastian Schneeweiss, Bre Robert J Glyn, Jun Liu, Seoyoung C Kim

Canagliflozin
GLP-1 RA
Placebo

Incidence of HF hospitalization
Months
Weeks

11
Secondary indications
Why did these database studies come to the same causal conclusion?
Cardiovascular outcomes associated with canagliflozin versus other non-gliflozin antidiabetic drugs: population based cohort study

Elisabetta Patorno, Allison B Goldfine, Sebastian Schneeweiss, Bret Robert J Glynn, Jun Liu, Seoyoung C Kim

How confident are we that the next study will get it right?
Re-analysis of Hemkens et al. BMJ 2016


<table>
<thead>
<tr>
<th>Study</th>
<th>ROR (95% CI)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holman 2000</td>
<td>0.74 (0.20, 2.68)</td>
<td>3.25</td>
</tr>
<tr>
<td>Shavelle 2002</td>
<td>0.96 (0.43, 2.13)</td>
<td>8.44</td>
</tr>
<tr>
<td>Winkelmayer 2002</td>
<td>0.55 (0.15, 2.02)</td>
<td>3.17</td>
</tr>
<tr>
<td>Karhik 2003</td>
<td>0.23 (0.02, 2.40)</td>
<td>0.99</td>
</tr>
<tr>
<td>Guru 2006</td>
<td>0.66 (0.06, 7.32)</td>
<td>0.93</td>
</tr>
<tr>
<td>Wu 2008</td>
<td>0.59 (0.12, 2.96)</td>
<td>2.09</td>
</tr>
<tr>
<td>Asclone 2003</td>
<td>0.28 (0.02, 4.43)</td>
<td>0.71</td>
</tr>
<tr>
<td>Polkinghorne 2004</td>
<td>0.79 (0.40, 1.56)</td>
<td>11.78</td>
</tr>
<tr>
<td>Gnerlich 2007</td>
<td>1.46 (0.88, 2.43)</td>
<td>20.90</td>
</tr>
<tr>
<td>Lindeauer 2004</td>
<td>1.01 (0.56, 1.82)</td>
<td>15.40</td>
</tr>
<tr>
<td>Butler 2009</td>
<td>0.53 (0.30, 0.95)</td>
<td>15.81</td>
</tr>
<tr>
<td>Cabell 2005</td>
<td>0.49 (0.03, 9.19)</td>
<td>0.63</td>
</tr>
<tr>
<td>Kim 2009</td>
<td>1.10 (0.30, 4.11)</td>
<td>3.11</td>
</tr>
<tr>
<td>Moss 2003</td>
<td>1.82 (0.80, 4.13)</td>
<td>7.98</td>
</tr>
<tr>
<td>Fonarow 2008</td>
<td>2.30 (0.65, 8.12)</td>
<td>3.38</td>
</tr>
<tr>
<td>Hahn 2010</td>
<td>4.53 (0.65, 31.43)</td>
<td>1.43</td>
</tr>
<tr>
<td>Overall (I-squared = 13.0%, p = 0.305)</td>
<td>0.98 (0.78, 1.23)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

© 2018 Harvard / Brigham Division of Pharmacoepidemiology
### Re-analysis of Hemkens et al. BMJ 2016


<table>
<thead>
<tr>
<th>Study</th>
<th>ROR (95% CI)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holman 2000</td>
<td>0.74 (0.20, 2.68)</td>
<td>3.25</td>
</tr>
<tr>
<td>Shavelle 2002</td>
<td>0.96 (0.43, 2.13)</td>
<td>8.44</td>
</tr>
<tr>
<td>Winkelmayer 2002</td>
<td>0.55 (0.15, 2.02)</td>
<td>3.17</td>
</tr>
<tr>
<td>Karhik 2003</td>
<td>0.23 (0.02, 2.40)</td>
<td>0.99</td>
</tr>
<tr>
<td>Guru 2006</td>
<td>0.66 (0.06, 7.32)</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Such summary statements do not inform us about the reasons of failure or success in a given study.
Confidence in validity of study findings

- Intended effect (confirmatory)
- Unintended effect (discovery)

Beneficial effect
- RCT
- RWD

Harmful effect
- RCT
- RWD
Confidence in validity of study findings

- **Canagliflozin and HF**
  - Unintended effect (discovery)
  - Intended effect (confirmatory)

- **Dabigatran and stroke**

- **Beneficial effect**
  - RCT
  - RWD

- **Harmful effect**
  - RCT
  - RWD

- **Aprotinin and death**
- **Tociluzimab and CVD**
- **SGLT-2 and DKA**
A spectrum of choices for decision makers

- Strongly prefer RCT
- RWD analysis possible

This talk
Reminder: Why we love RCTs

Randomized Controlled Trials

- Random treatment assignment
- Controlled outcome measurement
- Clear and easy to understand implementation
When to do database studies?

Study question-dependent

1. Active comparator preferred
2. Outcome, exposure measurable
3. Key confounders measurable
When to do database studies?

Study question-dependent:
1. Active comparator preferred
2. Outcome, exposure measurable
3. Key confounders measurable

The universe of study questions validly answerable

- RCTs
- RWD analyses

RCTs for regulatory decision making that could be replaced?
4. Proceed if
   a) Outcome observable with specificity
   b) Sufficient outcome surveillance
   c) Sufficient patient similarity is reached¹)

5. Avoid known design and analytic flaws:
   a) Avoid immortal time bias
   b) Avoid adjusting for causal intermediates
   c) Avoid reverse causation
   d) Deal with time-varying hazards

6. Do robustness checks
   a) Negative/positive controls
   b) Check balance of unmeasured factors

¹) Franklin et al. Epidemiology 2014
The advantages of an **active comparator new user design** has been demonstrated many times: Example Statin and mortality.

---

**Increasing Levels of Restriction in Pharmacoepidemiologic Database Studies of Elderly and Comparison With Randomized Trial Results**

Sebastian Schneeweiss, MD, ScD,* Amanda R. Patrick, MS,* Til Stürmer, MD, MPH,* M. Alan Brookhart, PhD,* Jerry Avorn, MD,* Malcolm Macleod, ScD,* Kenneth J. Rothman, DMD, DrPH,† and Robert J. Glynn, PhD, ScD*

---

- Increasing restriction of a broad RWD population leads to a narrow RCT population.
- The observed effect size is moving to the RCT finding with increasing restriction even w/o statistical adjustment.
4. Proceed if
   a) Outcome observable with specificity
   b) Sufficient outcome surveillance
   c) Sufficient patient similarity is reached

5. Avoid known design and analytic flaws:
   a) Avoid immortal time bias
   b) Avoid adjusting for causal intermediates
   c) Avoid reverse causation
   d) Deal with time-varying hazards

6. Do robustness checks
   a) Negative/positive controls
   b) Check balance of unmeasured factors

1) Franklin et al. Epidemiology 2014
Checking balance of unmeasured covariates in EHR-defined measures

Claims-defined
120 variables in 1:1 PS matching

EHR-defined
6 variables for balance checking
- Smoking
- BMI
- DM duration
- HbA1c
- eGFR
- LDL
Checking balance of unmeasured covariates in EHR-defined measures

Balance Analysis

<table>
<thead>
<tr>
<th></th>
<th>Linagliptin</th>
<th>Pioglitazone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never smoking</td>
<td>32.4%</td>
<td>33.9%</td>
</tr>
<tr>
<td>Obese</td>
<td>49.4%</td>
<td>46.1%</td>
</tr>
<tr>
<td>&gt;3 years DM duration</td>
<td>17.7%</td>
<td>20.1%</td>
</tr>
<tr>
<td>HbA1c, %</td>
<td>8.0 (7.1-9.1)</td>
<td>8.2 (7.1-9.9)</td>
</tr>
<tr>
<td>eGFR, ml/min/1.73m²</td>
<td>102 (93-116)</td>
<td>104 (96-118)</td>
</tr>
<tr>
<td>LDL, mg/dl</td>
<td>97 (73-116)</td>
<td>97 (79-115)</td>
</tr>
</tbody>
</table>

Sensitivity Analysis
A pathway


RCT No → RCT No → RCT No

Validated RWD analytics platform with audit trails
A pathway with regulatory validation

Is setting adequate for RWD analysis? Yes → No

Is data quality fit for purpose? Yes → No

Statistical analysis plan (ct.gov; encepp.eu) → Was balance achieved? Yes → No

Validation RWD analytics platform with audit trails

RCT No

Plan for additional analyses and checks

Regulatory and HTA consideration

Analysis → Structured reporting
A pathway with regulatory validation

**Is setting adequate for RWD analysis?**
- Yes
- No

**Is data quality fit for purpose?**
- Yes
- No

Statistical analysis plan (ct.gov; encepp.eu)

**Was balance achieved?**
- Yes
- No

**Plan for additional analyses and checks**

**Analysis**

**Structured reporting**

调节和HTA考虑

**Plan for additional analyses and checks**

© 2018 Harvard / Brigham Division of Pharmacoepidemiology
Telmisartan is an angiotensin receptor blocker (ARB)

Original indication in 1998:
• Hypertension

Supplementary indication in 2009:
• Cardiovascular risk reduction in patients ≥55 years

ONTARGET trial:
- Telmisartan (ARB) vs. Ramipril (ACE)
- CV death, MI, stroke, heart failure hospitalization
Case study: Telmisartan

- Let us say we have healthcare claims data available to us
- Let us say we have claims from commercial US insurer, e.g., MarketScan, from 2003 through 2009 (130 million lives covered).
Case study: Telmisartan

New user, active comparator, PS-matched cohort study

Table 1. Baseline characteristics prior to receiving telmisartan or ramipril

<table>
<thead>
<tr>
<th></th>
<th>Unmatched Population</th>
<th></th>
<th>PS-Matched Population</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ramipril (N=48,053)</td>
<td>Telmisartan (N=4665)</td>
<td>SD</td>
<td>Ramipril (N=4665)</td>
</tr>
<tr>
<td>Mean age (S. Dev.)</td>
<td>68.29 (9.52)</td>
<td>69.43 (9.60)</td>
<td>0.119</td>
<td>69.36 (9.67)</td>
</tr>
<tr>
<td>Age category</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55-60</td>
<td>9,747 (20.3%)</td>
<td>802 (17.2%)</td>
<td>0.149</td>
<td>839 (18.0%)</td>
</tr>
<tr>
<td>60-65</td>
<td>11,539 (24.0%)</td>
<td>985 (21.1%)</td>
<td></td>
<td>947 (20.3%)</td>
</tr>
<tr>
<td>65-70</td>
<td>6,262 (13.0%)</td>
<td>626 (13.4%)</td>
<td></td>
<td>655 (14.0%)</td>
</tr>
<tr>
<td>70-75</td>
<td>6,468 (13.5%)</td>
<td>681 (14.6%)</td>
<td></td>
<td>666 (14.3%)</td>
</tr>
<tr>
<td>≥75</td>
<td>14,037 (29.2%)</td>
<td>1,571 (33.7%)</td>
<td></td>
<td>1,558 (33.4%)</td>
</tr>
<tr>
<td>Male</td>
<td>31,940 (66.5%)</td>
<td>2,413 (51.7%)</td>
<td>0.303</td>
<td>2,343 (50.2%)</td>
</tr>
<tr>
<td>Date of cohort entry</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First Quarter</td>
<td>13,667 (28.4%)</td>
<td>1,198 (25.7%)</td>
<td>0.046</td>
<td>1,149 (24.6%)</td>
</tr>
<tr>
<td>Second Quarter</td>
<td>10,080 (21.0%)</td>
<td>1,038 (22.3%)</td>
<td></td>
<td>1,005 (21.5%)</td>
</tr>
<tr>
<td>Third Quarter</td>
<td>12,730 (26.5%)</td>
<td>1,310 (28.1%)</td>
<td></td>
<td>1,395 (29.9%)</td>
</tr>
<tr>
<td>Fourth Quarter</td>
<td>11,576 (24.1%)</td>
<td>1,119 (24.0%)</td>
<td></td>
<td>1,116 (23.9%)</td>
</tr>
</tbody>
</table>
Case study: Telmisartan

Balanced patient characteristics after PS-matching

<table>
<thead>
<tr>
<th>Comorbid Conditions</th>
<th>Group 1</th>
<th>Group 2</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension</td>
<td>21,361</td>
<td>2,835</td>
<td>0.331</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>37,591</td>
<td>3,105</td>
<td>0.263</td>
</tr>
<tr>
<td>Diabetes Mellitus</td>
<td>14,375</td>
<td>1,524</td>
<td>0.059</td>
</tr>
<tr>
<td>PAD</td>
<td>2,651</td>
<td>362</td>
<td>0.09</td>
</tr>
<tr>
<td>Stroke or TIA</td>
<td>5,727</td>
<td>730</td>
<td>0.108</td>
</tr>
<tr>
<td>Angina</td>
<td>11,272</td>
<td>815</td>
<td>0.149</td>
</tr>
<tr>
<td>Heart failure</td>
<td>7,205</td>
<td>510</td>
<td>0.121</td>
</tr>
<tr>
<td>Renal disease</td>
<td>3,549</td>
<td>545</td>
<td>0.147</td>
</tr>
<tr>
<td>Smoking</td>
<td>1,734</td>
<td>115</td>
<td>0.067</td>
</tr>
<tr>
<td>Previous CABG or PCI</td>
<td>5,454</td>
<td>124</td>
<td>0.346</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Medications</th>
<th>Group 1</th>
<th>Group 2</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statin</td>
<td>22,441</td>
<td>2,104</td>
<td>0.032</td>
</tr>
<tr>
<td>Beta-Blocker</td>
<td>20,957</td>
<td>1,926</td>
<td>0.047</td>
</tr>
<tr>
<td>Anti-platelet agent</td>
<td>11,031</td>
<td>1,127</td>
<td>0.028</td>
</tr>
<tr>
<td>Calcium-channel blocker</td>
<td>5,386</td>
<td>833</td>
<td>0.189</td>
</tr>
<tr>
<td>Diuretic</td>
<td>11,396</td>
<td>1,342</td>
<td>0.115</td>
</tr>
<tr>
<td>ACE or ARB use</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
# Case study: Telmisartan

## Comparing RWE vs. RCT results

<table>
<thead>
<tr>
<th></th>
<th>Observational Cohort Study</th>
<th>ONTARGET Clinical Trial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ramipril (N=4,665)</td>
<td>Telmisartan (N=4,665)</td>
</tr>
<tr>
<td><strong>Composite endpoint</strong></td>
<td>Ref. 0.99 (0.85, 1.14)*</td>
<td>1.01 (0.94, 1.09)</td>
</tr>
<tr>
<td><strong>Stroke</strong></td>
<td>Ref. 0.95 (0.71, 1.26)*</td>
<td>0.91 (0.70, 1.05)</td>
</tr>
<tr>
<td><strong>Myocardial infarction</strong></td>
<td>Ref. 0.92 (0.67, 1.27)*</td>
<td>1.07 (0.94, 1.22)</td>
</tr>
<tr>
<td><strong>Hospitalization for heart failure</strong></td>
<td>Ref. 0.95 (0.79, 1.13)*</td>
<td>1.12 (0.97, 1.29)</td>
</tr>
<tr>
<td><strong>Angioedema</strong></td>
<td>Ref. 0.13 (0.03, 0.56)*</td>
<td>0.4 (p=0.01)**</td>
</tr>
</tbody>
</table>
### Case study: Telmisartan

**Comparing RWE vs. RCT results**

<table>
<thead>
<tr>
<th></th>
<th>Observational Cohort Study</th>
<th>ONTARGET Clinical Trial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ramipril (N=4,665)</td>
<td>Telmisartan (N=4,665)</td>
</tr>
<tr>
<td><strong>Composite endpoint</strong></td>
<td>Ref.</td>
<td>0.99 (0.85, 1.14)*</td>
</tr>
<tr>
<td><strong>Stroke</strong></td>
<td>Ref.</td>
<td>0.95 (0.71, 1.26)*</td>
</tr>
<tr>
<td><strong>Myocardial infarction</strong></td>
<td>Ref.</td>
<td>0.92 (0.67, 1.27)*</td>
</tr>
<tr>
<td><strong>Hospitalization for heart failure</strong></td>
<td>Ref.</td>
<td>0.95 (0.79, 1.13)*</td>
</tr>
<tr>
<td><strong>Angioedema</strong></td>
<td>Ref.</td>
<td>0.13 (0.03, 0.56)*</td>
</tr>
</tbody>
</table>
Transparency to increase confidence in RWD analyses

Randomized Controlled Trials

- Random treatment assignment
- Controlled outcome measurement
- Clear and easy to understand implementation

Non-interventional Database Studies

- Study design choices balance patient characteristics
- Non-standardized observations
- Complex study design and analytic methods

Controlled study environment and self-evident methodology provides confidence in decision making.

Transparent, structured reporting of complex methodology clarifies study validity for decision makers.
How to ...

7. Use validated RWE software platform
   a) Avoids design flaws
   b) Increased transparency
   c) Stores audit trails

1) Franklin et al. Epidemiology 2014
Structured user interfaces guide the user through the process collecting all study parameters

Select patients in transparent and reproducible ways

Select risk adjustment

Select comparison group

Select follow-up model
Analytic tools are build for 100% transparency

Tabular format (FDA Sentinel)

<table>
<thead>
<tr>
<th>Exposure of Interest</th>
<th>Comparator of Interest</th>
<th>Enrolment Gap: 95 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glibenclamide, glimepiride, and other second-generation sulfonylureas (e.g. glyburide, glimepiride, and other second-generation sulfonylureas)</td>
<td>Glipizide, glyburide, and other second-generation sulfonylureas (e.g. glyburide, glimepiride, and other second-generation sulfonylureas)</td>
<td>188</td>
</tr>
<tr>
<td>Incident: Hospitalization</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Incident: Death</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Death: 30</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Death: 60</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Death: 90</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>Death: 120</td>
<td>120</td>
<td></td>
</tr>
</tbody>
</table>

Text and tabular format

Preprocedure score analysis

This section describes the general approach to preprocedure score analysis used throughout the analysis. Specific uses of the techniques described below are indicated in later sections.

Preprocedure score analysis: Propensity scores were estimated using logistic regression. The treatment was specified as the dependent variable. All covariates listed were entered as independent variables, and for each variable adjustment was made. The probability score was predicted using logistic regression models. A prematching model contains covariates used in the final covariate list. Preprocedure score analysis is used to predict whether the treatment will be continued.

Preprocedure score analysis: Propensity scores were divided into deciles and indicated the decile of PS were entered into the outcome model, alongside exposure and basic covariates. The PS decile was used as the reference category.

Preprocedure score analysis: Preprocedure score analysis was performed. The PS matching, PS matching, PS was chosen to reduce confounding by eliminating patients with highly improbable treatment choices who appear in the estimates of the PS distributions. Among the exposed patients, the PS distribution of the propensity score was determined; any patient whose propensity score exceeded this value was removed from the analysis. Similarly, among the reference group patients, the 2.5% smallest of the propensity score distribution was removed from the analysis. Deciles of the propensity scores were determined from the remaining values, and each patient was assigned an index of decile of propensity scores. The propensity score indicated the decile of PS were entered into the outcome model, alongside exposure and basic covariates. The PS decile was used as the reference category. More after triming, the study population was at least 95% similar to the pre-treatment populations.

Preprocedure score matched analysis

Propensity score matching was performed using 1:1 nearest neighbor matching with a maximum matching cutoff of 0.20. A matched propensity score analysis, multivariate adjustment was achieved through the matching process. After matching, the treatment effect measure were then derived from the balanced populations without any further adjustment. We also explored specific propensity score distributions imported into the treatment calculation. Differences in the estimator distributions between exposure groups are displayed to assess whether differences in exposure are consistent with the treatment effect estimates. The PS matching is compared as a measure of balance for covariate balance. Covariates close to 0.5 represent goal oriented balance.

Statistical Analysis

Primary Analysis

Covariate assessment

In the primary analysis, covariates were examined in the PS matching prior to entry. The index day was included in the covariate assessment period.
Line programming against line programming (double programming, same protocol):
- 10+ examples at BWH
- In 100% get different findings!

Line programming against FDA Sentinel tools:
- 3 examples at BWH
- In 100% there were errors in line programming

Line programming against Aetion platform:
- More than 50 validation activities in >20 organizations
- In 100% of activities there were errors in line programming
  - 70% misinterpretation or alternative assumption
  - 30% coding errors (time related, definitions)

⇒ Line programming for healthcare database analytics is
  1) Inherently error prone
  2) Not validatable at scale
Sharing programming code is not helpful...  

... as it does not clarify whether the indented study was implemented accurately

3) Lacks transparency
4) Lacks reproducibility against intended protocol
Lack of reporting details make RWD studies non-reproducible
Quantify the current state of reproducibility of database studies

- Systematic search using Google Scholar
- Include descriptive or comparative safety/effectiveness cohort studies
- Randomly sample 250 studies
- Consider all publically available information
- Replicate studies (blind to original results)
- Contact original authors to understand assumptions, differences
- Top h-5 clinical, epidemiology journals
  - Published after Jan 1, 2011
  - “cohort” + “claims” + database name
- CONSORT style diagram
  - Exclude if data source mismatch or PDF unavailable
    - Standardized extraction form
- Metrics to quantify reproducibility
  - Frequency insufficient reporting, Std. Diff, etc.
# Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0

Shirley V. Wang\(^1\)\(^2\)  |  Sebastian Schneeweiss\(^1\)\(^2\)  |  Marc L. Berger\(^3\)  |  Jeffrey Brown\(^4\)  
Frank de Vries\(^5\)  |  Ian Douglas\(^6\)  |  Joshua J. Gagne\(^1\)\(^2\)  |  Rosa Gin\(^7\)  |  Olaf Klungel\(^8\)  
C. Daniel Mullins\(^9\)  |  Michael D. Nguyen\(^10\)  |  Jeremy A. Rassen\(^11\)  |  Liam Smeeth\(^6\)  |  Miriam Sturkenboom\(^12\)  
on behalf of the joint ISPE-ISPOR Special Task Force on Real World Evidence in Health Care Decision Making

---

**TABLE 2** Reporting specific parameters to increase reproducibility of database studies

<table>
<thead>
<tr>
<th>Description</th>
<th>Example</th>
<th>Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Reporting on data source should include:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A.1 Data provider</td>
<td>Data source name and name of organization that provided data.</td>
<td>Medicaid Analytic Extracts data covering 50</td>
</tr>
<tr>
<td><strong>A.2 Data extraction date (DED)</strong></td>
<td>The date (or version number) when the data were extracted from the dynamic raw data stream. Use the date that the data were extracted for research use by the vendor.</td>
<td></td>
</tr>
<tr>
<td>A.3 Data sampling</td>
<td>The search/extraction criteria applied to source data accessible to the researcher and the subset of the data available from this data.</td>
<td></td>
</tr>
<tr>
<td>A.4 Source data range (SDR)</td>
<td>The calendar time range of data used in the study. Note that the implementation of the time range only use a subset of the available data.</td>
<td></td>
</tr>
</tbody>
</table>

**D. Reporting on exposure definition should include:**

| D.1 Type of exposure | The type of exposure that is captured or measured, e.g., drug versus procedure, new use, incident, prevalent, cumulative, time-varying. | We evaluated risk of outcome Z following incident exposure to drug X or drug Y. Incident exposure was defined as beginning on the day of the first dispensation for one of these drugs after at least 180 days without dispensations for either (SED). Patients with incident exposure to both drug X and drug Y on the same SED were excluded. The exposure risk window for patients with Drug X and Drug Y began 10 days after incident exposure and continued until 14 days past the last days supply, including refills. If a patient refilled early, the date of the early refill and subsequent refills were adjusted so that the full days supply from the initial dispensation was counted before the days supply from the next dispensation was tallied. Gaps of less than or equal to 14 days between one dispensation plus days supply and the next dispensation for the same drug were bridged (i.e., the time was | |
| D.2 Exposure risk window (ERW) | The ERW is specific to an exposure and the outcome under investigation. For drug exposures, it is equivalent to the time between the minimum and maximum hypothesized induction time following ingestion of the molecule. | Drug era, risk window | |
| D.2a Induction period\(^1\) | Days on or following study entry date during which an outcome would not be counted as "exposed time" or "comparator time". | Blackout period | |
| D.2b Stockpiling\(^1\) | The algorithm applied to handle leftover days supply if there are early refills. | Episode gap, grace period, persistence window, gap days | |
| D.2c Bridging exposure episodes\(^1\) | The algorithm applied to handle gaps that are longer than expected if there was perfect adherence (e.g., non-overlapping dispensation + days supply). | | |

---

\(^1\) Note: Values \([^1-4]_1\) are not necessarily mutually exclusive.
How well can RWD analyses reproduce RCT findings?

**Process**

1. **Candidate RCTs**
2. **Select target RCTs**
3. **Set up scalable RWD analytics platform**
4. **Reproduce RCTs with RWD**

**Products**

1. **List of RCTs to be reproduced with RWD**
2. **Document exclusions:**
   - Limited RWD, Key measurements missing,
   - Extremely strong confounding etc. ...
3. **RWD study infrastructure:**
4. **Scalable RWD infrastructure**
5. **Quantify accuracy of RWD studies**
6. **Expert group guidance**
RWE fit for Decision Making in Healthcare

MVET framework for RWE that is fit for DM

CP&T 2016;100:633-46

ISPE/ISPOR consensus paper on reproducibility

Pharmacoepi Drug Saf 2017;9:1018-32

When and how to augment RCTs with RWE

CP&T 2017;102:924-33

Real World Data in Adaptive Biomedical Innovation: A Framework for Generating Evidence Fit for Decision-Making


Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0

Shirley V. Wang1,2 | Sebastian Schneeweiss1,2 | Marc L. Berger3 | Jeffrey Brown4 | Frank de Vries5 | Ian Douglas6 | Joshua J. Gagne1,2 | Rosa Gini7 | Olaf Klungel8 | C. Daniel Mullins9 | Michael D. Nguyen10 | Jeremy A. Rassen11 | Liam Smeeth4 | Miriam Sturkenboom12

on behalf of the joint ISPE-ISPOR Special Task Force on Real World Evidence in Healthcare Decision Making

When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials?

Jessica M. Franklin1 and Sebastian Schneeweiss1
A pathway

Is setting adequate for RWD analysis? Yes

Is data quality fit for purpose? Yes

Statistical analysis plan (ct.gov; encepp.eu)

Was balance achieved? Yes

Analysis

Validated RWD analytics platform with audit trails

Plan for additional analyses and checks

Regulatory and HTA consideration

Structured reporting

RCT

No

RCT

No

RCT

No