Large-Scale Evidence Generation and Evaluation in a Network of Databases (LEGEND)

Patrick Ryan, Martijn Schuemie, Marc Suchard
on behalf of the LEGEND team

OHDSI Symposium
12 October 2018
Association is not causation: treatment effects cannot be estimated from observational data in heart failure

Christopher J. Rush, Ross T. Campbell, Pardeep S. Jhund, Mark C. Petrie, and John J.V. McMurray*

British Heart Foundation Cardiovascular Research Centre, Institute of Cardiovascular and Medical Sciences, University of Glasgow, 126 University Place, Glasgow G12 8TA, UK

Received 16 January 2018; revised 1 April 2018; editorial decision 22 June 2018; accepted 27 June 2018; online publish-ahead-of-print 1 August 2018

Aims
Treatment ‘effects’ are often inferred from non-randomized and observational studies. These studies have inherent biases and limitations, which may make therapeutic inferences based on their results unreliable. We compared the conflicting findings of these studies to those of prospective randomized controlled trials (RCTs) in relation to pharmacological treatments for heart failure (HF).

Methods and results
We searched Medline and Embase to identify studies of the association between non-randomized drug therapy and all-cause mortality in patients with HF until 31 December 2017. The treatments of interest were: angiotensin-
Residual study bias

Rush et al., 2018
Published observational study results

Suspicious cutoff at p=0.05
- Publication bias (leads to false positives)
- P-hacking (leads to false positives)

59,196 estimates
19.0% of CIs include 1
Trouble with observational research

• Individual studies are often biased due to *confounding, selection bias, and measurement error*

• Across studies, observational research as a whole is even more biased due to *publication bias* and *p-hacking*
Improving methods to address confounding

• Construct large generic set of covariates
  – $10,000 < n < 100,000$

• Use regularized regression to fit propensity model

• Match or stratify on propensity score

Achieving balance on all 58,285 covariates
Measuring residual bias

Control questions:

- exposure-outcome pairs with known effect size
- negative and positive controls

Empirical calibration:

- Adjust p-value and confidence interval using estimates for controls
Solving publication bias and p-hacking

• Fully specified protocols
• Pre-registering studies
• Open science
• Large-scale studies...
Depression proof of concept

Define a large set of related research questions

- Indication: depression
- Treatments: n = 17
- Treatments pairs: n = 272
- Outcomes: n = 22

Research questions: n = 5,984

Select negative controls & synthesize positive controls

- Negative controls: n = 52 per treatment pair
- Positive controls: n = 156 per treatment pair

Control questions: n = 56,266

Systematic observational research process

- Select study population
  - e.g. excluding subjects with prior outcome
- Create propensity scores
- Stratify by propensity scores
- Fit outcome (survival) model

Effect size estimates: n = 5,984

Control estimates: n = 56,266

Calibration models: n = 272

Calibrated estimates: n = 5,984

Evaluate & calibrate
Results of proposed solution

- Includes information on small effect sizes
- Study bias assessed using negative and positive controls
- No p-hacking, no publication bias
Depression results publicly available

http://data.ohdsi.org/SystematicEvidence/
LEGEND

LARGE-SCALE EVIDENCE GENERATION AND EVALUATION IN A NETWORK OF DATABASES
Building the process to generate the evidence
1. Evidence will be generated at **large-scale**.

2. **Dissemination** of the evidence will not depend on the estimated effects.

3. Evidence will be generated by consistently applying a **systematic approach** across all research questions.

4. The evidence will be generated using a **pre-specified** analysis design.

5. The evidence will be generated using **open source** software that is freely available to all.

6. The evidence generation process will be **empirically evaluated** by including control research questions where the true effect size is known.

7. The evidence will be generated using **best-practices**.

8. LEGEND will **not** be used to **evaluate methods**.

9. The evidence will be **updated** on a regular basis.

10. **No patient-level data** will be shared between sites in the network, only aggregated data.
1. Evidence will be generated at **large-scale**.

2. **Dissemination** of the evidence will not depend on the estimated effects.

3. Evidence will be generated by consistently applying a **systematic approach** across all research questions.

4. The evidence will be generated using a **pre-specified** analysis design.

5. The evidence will be generated using **open source** software that is freely available to all.

6. The evidence generation process will be **empirically evaluated** by including control research questions where the true effect size is known.

7. The evidence will be generated using **best-practices**.

8. LEGEND will **not** be used to **evaluate methods**.

9. The evidence will be **updated** on a regular basis.

10. **No patient-level data** will be shared between sites in the network, only aggregated data.
LEGEND Guiding Principles

1. Evidence will be generated at large-scale.
2. **Dissemination** of the evidence will not depend on the estimated effects.
3. Evidence will be generated by consistently applying a **systematic approach** across all research questions.
4. The evidence will be generated using a **pre-specified** analysis design.
5. **The evidence will be generated using open source software that is freely available to all.**
6. The evidence generation process will be **empirically evaluated** by including control research questions where the true effect size is known.
7. The evidence will be generated using **best-practices**.
8. LEGEND will **not** be used to **evaluate methods**.
9. The evidence will be **updated** on a regular basis.
10. **No patient-level data** will be shared between sites in the network, only aggregated data.
Evidence generation

- Research questions
- Methods
- Databases

Evidence base
Research questions

Previously: Depression treatments
This run: Hypertension treatments
‘Target trial’ to compare two initial therapies

**Treatment strategies:**
- Monotherapy with ACE
- Monotherapy with ARB

**Eligibility criteria:**
- Diagnosed with hypertension in 1 year prior to index
- No prior antihypertensive drug use anytime prior to index

**Causal contrasts of interest:**
- Intent-to-treat effect
- On-treatment effect

**Outcomes:**
- **Efficacy:**
  - Myocardial infarction
  - Stroke
  - Heart Failure
- **Safety:**
  - Known or potential adverse events, e.g.
    - Acute renal failure
    - Angioedema
    - Cough
    - Diarrhea
    - Fall
    - Gout
    - Headache
    - Hyperkalemia
    - Hyponatremia
    - Hypotension
    - Impotence
    - Syncope
    - Vertigo

**Index:**
- Time zero

**Analysis plan:**
- Time-to-first-event analysis
- Cox proportional hazards

**Follow-up time**

**Medical history lookback time**

**Randomization**
Observational study to compare two initial therapies

**Eligibility criteria:**
- Diagnosed with hypertension in 1 year prior to index
- No prior antihypertensive drug use anytime prior to index

**Treatment strategies:**
- Monotherapy with ACE
- Monotherapy with ARB

**Causal contrasts of interest:**
- Intent-to-treat effect
- On-treatment effect

**Outcomes:**
- Efficacy:
  - Myocardial infarction
  - Stroke
  - Heart Failure
- Safety:
  - Known or potential adverse events, e.g.
    - Acute renal failure
    - Angioedema
    - Cough
    - Diarrhea
    - Fall
    - Gout
    - Headache
    - Hyperkalemia
    - Hyponatremia
    - Hypotension
    - Impotence
    - Syncope
    - Vertigo

**Analysis plan:**
- Time-to-first-event analysis
- Cox proportional hazards
Hypertension mono-therapy

Duo-therapy

Truven Health MarketScan CCAE. Therapies > 2 ingredients not shown
Comparisons of hypertension treatments

<table>
<thead>
<tr>
<th></th>
<th>Theoretical</th>
<th>Observed (n &gt; 2,500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single ingredients</td>
<td>58</td>
<td>39</td>
</tr>
<tr>
<td>Single ingredient comparisons</td>
<td>58 * 57 = 3,306</td>
<td>1,296</td>
</tr>
<tr>
<td>Single drug classes</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>Single class comparisons</td>
<td>15 * 14 = 210</td>
<td>156</td>
</tr>
</tbody>
</table>
‘Target trial’ to compare mono vs combination therapy

**Treatment strategies:**
- Monotherapy with ACE
- Monotherapy with THZ
- Combination therapy with ACE+THZ

**Eligibility criteria:**
- Diagnosed with hypertension in 1 year prior to index
- No prior antihypertensive drug use anytime prior to index

**Causal contrasts of interest:**
- Intent-to-treat effect
- On-treatment effect

**Outcomes:**
- **Efficacy:**
  - Myocardial infarction
  - Stroke
  - Heart Failure
- **Safety:**
  - Known or potential adverse events, e.g.
    - Acute renal failure
    - Angioedema
    - Cough
    - Diarrhea
    - Fall
    - Gout
    - Headache
    - Hyperkalemia
    - Hyponatremia
    - Hypotension

**Analysis plan:**
- Time-to-first-event analysis
- Cox proportional hazards
- Two pairwise comparisons:
  - ACE vs. ACE+THZ
  - THZ vs. ACE+THZ

**Index:**
- Time zero
Observational study to compare mono vs combination therapy

**Treatment strategies:**
- Monotherapy with ACE
- Monotherapy with THZ
- Combination therapy with ACE+THZ

**Eligibility criteria:**
- Diagnosed with hypertension in 1 year prior to index
- No prior antihypertensive drug use anytime prior to index

**Causal contrasts of interest:**
- Intent-to-treat effect
- On-treatment effect

**Outcomes:**
- **Efficacy:**
  - Myocardial infarction
  - Stroke
  - Heart Failure
- **Safety:**
  - Known or potential adverse events, e.g.
    - Acute renal failure
    - Angioedema
    - Cough
    - Diarrhea
    - Fall
    - Gout
    - Headache
    - Hyperkalemia
    - Hyponatremia
    - Hypotension
    - Impotence
    - Syncope
    - Vertigo

**Analysis plan:**
- Time-to-first-event analysis
- Cox proportional hazards
- Two pairwise comparisons:
  - ACE vs. ACE+THZ
  - THZ vs. ACE+THZ
Hypertension duo-therapy

Truven Health MarketScan CCAE. Therapies > 2 ingredients not shown
Comparisons of hypertension treatments

<table>
<thead>
<tr>
<th></th>
<th>Theoretical</th>
<th>Observed (n &gt; 2,500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single ingredients</td>
<td>58</td>
<td>39</td>
</tr>
<tr>
<td>Single ingredient comparisons</td>
<td>58 * 57 = 3,306</td>
<td>1,296</td>
</tr>
<tr>
<td>Single drug classes</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>Single class comparisons</td>
<td>15 * 14 = 210</td>
<td>156</td>
</tr>
<tr>
<td>Dual ingredients</td>
<td>58 * 57 / 2 = 1,653</td>
<td>58</td>
</tr>
<tr>
<td>Single vs duo drug comparisons</td>
<td>58 * 1,653 = 95,874</td>
<td>3,810</td>
</tr>
<tr>
<td>Dual classes</td>
<td>15 * 14 / 2 = 105</td>
<td>32</td>
</tr>
<tr>
<td>Single vs duo class comparisons</td>
<td>15 * 105 = 1,575</td>
<td>832</td>
</tr>
</tbody>
</table>
## Comparisons of hypertension treatments

<table>
<thead>
<tr>
<th></th>
<th>Theoretical</th>
<th>Observed (n &gt; 2,500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single ingredients</td>
<td>58</td>
<td>39</td>
</tr>
<tr>
<td>Single ingredient comparisons</td>
<td>58 * 57 = 3,306</td>
<td>1,296</td>
</tr>
<tr>
<td>Single drug classes</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>Single class comparisons</td>
<td>15 * 14 = 210</td>
<td>156</td>
</tr>
<tr>
<td>Dual ingredients</td>
<td>58 * 57 / 2 = 1,653</td>
<td>58</td>
</tr>
<tr>
<td>Single vs duo drug comparisons</td>
<td>58 * 1,653 = 95,874</td>
<td>3,810</td>
</tr>
<tr>
<td>Dual classes</td>
<td>15 * 14 / 2 = 105</td>
<td>32</td>
</tr>
<tr>
<td>Single vs duo class comparisons</td>
<td>15 * 105 = 1,575</td>
<td>832</td>
</tr>
<tr>
<td>Duo vs duo drug comparisons</td>
<td>1,653 * 1,652 = 2,730,756</td>
<td>2,784</td>
</tr>
<tr>
<td>Duo vs duo class comparisons</td>
<td>105 * 104 = 10,920</td>
<td>992</td>
</tr>
</tbody>
</table>
Comparisons of hypertension treatments

<table>
<thead>
<tr>
<th></th>
<th>Theoretical</th>
<th>Observed (n &gt; 2,500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single ingredients</td>
<td>58</td>
<td>39</td>
</tr>
<tr>
<td>Single ingredient comparisons</td>
<td>58 * 57 = 3,306</td>
<td>1,296</td>
</tr>
<tr>
<td>Single drug classes</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>Single class comparisons</td>
<td>15 * 14 = 210</td>
<td>156</td>
</tr>
<tr>
<td>Dual ingredients</td>
<td>58 * 57 / 2 = 1,653</td>
<td>58</td>
</tr>
<tr>
<td>Single vs duo drug comparisons</td>
<td>58 * 1,653 = 95,874</td>
<td>3,810</td>
</tr>
<tr>
<td>Dual classes</td>
<td>15 * 14 / 2 = 105</td>
<td>32</td>
</tr>
<tr>
<td>Single vs duo class comparisons</td>
<td>15 * 105 = 1,575</td>
<td>832</td>
</tr>
<tr>
<td>Duo vs duo drug comparisons</td>
<td>1,653 * 1,652 = 2,730,756</td>
<td>2,784</td>
</tr>
<tr>
<td>Duo vs duo class comparisons</td>
<td>105 * 104 = 10,920</td>
<td>992</td>
</tr>
<tr>
<td>Total comparisons</td>
<td>2,843,250</td>
<td>10,278</td>
</tr>
</tbody>
</table>
Not all comparisons are valid
Not all comparisons are valid
Not all comparisons are valid
Not all comparisons are valid
58 outcomes of interest

Abdominal pain
Abnormal weight gain
Abnormal weight loss
Acute myocardial infarction
Acute pancreatitis
Acute renal failure
All-cause mortality
Anaphylactoid reaction
Anemia
Angioedema
Anxiety
Bradycardia
Cardiac arrhythmia
Cardiovascular disease
Cardiovascular-related mortality
Chest pain or angina
Chronic kidney disease
Coronary heart disease
Cough
Decreased libido
Demenitia
Depression
Diarrhea
Edema
End stage renal disease
Fall
Gastrointestinal bleeding
Gout
Headache
Heart failure
Hemorrhagic stroke
Hepatic failure
Hospitalization with heart failure
Hospitalization with preinfarction syndrome
Hyperkalemia
Hypokalemia
Hypomagnesemia
Hyponatremia
Hypotension
Impotence
Ischemic stroke
Kidney disease
Malignant neoplasm
Measured renal dysfunction
Nausea
Neutropenia or agranulocytosis
Rash
Rhabdomyolysis
Stroke
Sudden cardiac death
Syncope
Thrombocytopenia
Transient ischemic attack
Type 2 diabetes mellitus
Vasculitis
Venous thromboembolic events
Vertigo
Vomiting
58 outcomes of interest

| Abdominal pain | Dementia | Ischemic stroke |
| Abnormal weight gain | Depression | Kidney disease |
| Abnormal weight loss | Diarrhea | Malignant neoplasm |
| Acute myocardial infarction | Edema | Measured renal dysfunction |
| Acute pancreatitis | End stage renal disease | Nausea |
| Acute renal failure | Fall | Neutropenia or agranulocytosis |
| All-cause mortality | Gastrointestinal bleeding | Rash |

<table>
<thead>
<tr>
<th>Theoretical</th>
<th>Observed (n &gt; 2,500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcomes of interest</td>
<td>58</td>
</tr>
<tr>
<td>Target-comparator-outcomes</td>
<td>2,843,250 * 58 = 164,908,500</td>
</tr>
</tbody>
</table>

- Bradycardia
- Cardiac arrhythmia
- Cardiovascular disease
- Cardiovascular-related mortality
- Chest pain or angina
- Chronic kidney disease
- Coronary heart disease
- Cough
- Decreased libido
- Hepatic failure
- Hospitalization with heart failure
- Hospitalization with preinfarction syndrome
- Hyperkalemia
- Hypokalemia
- Hypomagnesemia
- Hyponatremia
- Hypotension
- Impotence
- Thrombocytopenia
- Transient ischemic attack
- Type 2 diabetes mellitus
- Vasculitis
- Venous thromboembolic events
- Vertigo
- Vomiting
Each research question requires

- Evaluation of the propensity score distribution
- Evaluation of covariate balance
- Inclusion of negative and positive controls
- Empirical calibration
Abnormal cervical smear
Abnormal pupil
Abrasion and/or friction burn of trunk without infection
Absence of breast
Absent kidney
Acid reflux
Acquired hallux valgus
Acquired keratoderma
Acquired trigger finger
Acute conjunctivitis
Amputated foot
Anal and rectal polyp
Burn of forearm
Calcaneal spur
Cannabis abuse
Cervical somatic dysfunction
Changes in skin texture
Chondromalacia of patella
Cocaine abuse
Colostomy present
Complication due to Crohn's disease
Contact dermatitis
Contusion of knee
Crohn's disease
Derangement of knee
Difficulty sleeping
Disproportion of reconstructed breast
Effects of hunger
Endometriosis
Epidermoid cyst
Feces contents abnormal
Foreign body in orifice
Ganglion cyst
Genetic predisposition
Hammer toe
Hereditary thrombophilia
Herpes zoster without complication
High risk sexual behavior
Homocystinuria
Human papilloma virus infection
Ileostomy present
Impacted cerumen
Impingement syndrome of shoulder region
Ingrowing nail
Injury of knee
Irregular periods
Kwashiorkor
Late effect of contusion
Late effect of motor vehicle accident
Leukorrhea
Macular drusen
Melena
Nicotine dependence
Noise effects on inner ear
Non-specific tuberculin test reaction
Non-toxic multinodular goiter
Onychomycosis due to dermatophyte
Opioid abuse
Passing flatus
Postviral fatigue syndrome
Presbyopia
Problem related to lifestyle
Psychalgia
Ptotic breast
Regular astigmatism
Senile hyperkeratosis
Somatic dysfunction of lumbar region
Splinter of face, without major open wound
Sprain of ankle
Strain of rotator cuff capsule
Tear film insufficiency
Tobacco dependence syndrome
Vaginitis and vulvovaginitis
Verruca vulgaris
Wrist joint pain
Wristdrop
76 negative controls

<table>
<thead>
<tr>
<th>Condition</th>
<th>Theoretical</th>
<th>Observed (n &gt; 2,500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative control outcomes</td>
<td>76</td>
<td>76</td>
</tr>
<tr>
<td>Target-comparator-neg controls</td>
<td>2,843,250 * 76 = 216,087,000</td>
<td>769,476</td>
</tr>
<tr>
<td>Positive control outcomes</td>
<td>76 * 3 = 228</td>
<td>228</td>
</tr>
<tr>
<td>Target-comparator-pos controls</td>
<td>2,843,250 * 228 = 648,261,000</td>
<td>662,484</td>
</tr>
<tr>
<td>Total control target-comparator-outcomes</td>
<td>864,348,000</td>
<td>1,431,960</td>
</tr>
</tbody>
</table>

- Abnormal cervical smear
- Abnormal pupil
- Abrasion and/or friction burn of trunk without infection
- Absence of breast
- Absent kidney
- Acid reflux
- Acquired hallux valgus
- Acquired keratoderma
- Acquired trigger finger
- Disproportion of reconstructed breast
- Effects of hunger
- Endometriosis
- Epidermoid cyst
- Feces contents abnormal
- Foreign body in orifice
- Ganglion cyst
- Genetic predisposition
- Hammer toe
- Nicotine dependence
- Noise effects on inner ear
- Non-specific tuberculin test reaction
- Non-toxic multinodular goiter
- Onychomycosis due to dermatophyte
- Opioid abuse
- Passing flatus
- Postviral fatigue syndrome
- Presbyopia

Additional conditions:
- Contusion of knee
- Crohn's disease
- Derangement of knee
- Difficulty sleeping
- Late effect of motor vehicle accident
- Leukorrhea
- Macular drusen
- Melena
- Wrist joint pain
- Wristdrop
Methods

This run:
• Emulate target trial: new-user cohort design
• Expert-crafted outcome definitions
• Large scale propensity models
• Stratification + variable ratio matching
• Empirical calibration

Not static. Driven by **defined best practices**, driven by **empirical evaluation**
Databases

Previously: 4 US insurance databases

This run:

- **US insurance databases**
  - IBM® MarketScan® CCAE
  - IBM® MarketScan® MDCD
  - IBM® MarketScan® MDCR
  - Optum© Clininformatics
- **Japanese insurance database**
  - Japan Medical Data Center
- **Korean national insurance database**
  - NHIS-NSC
- **US EHR databases**
  - Columbia University Medical Center
  - Optum© PANTHER®
- **German EHR database**
  - QuintilesIMS Disease Analyzer (DA) Germany
LEGEND results

- 1,321,696 estimates
- 83.4% of CIs includes 1

Graph showing hazard ratio on the x-axis and standard error on the y-axis.
Dissemination

Evidence generation

- Research questions
- Methods
- Databases

Evidence base

- Web app
- Papers
- 3rd parties
http://data.ohdsi.org/LegendBasicViewer/
http://data.ohdsi.org/LegendMedCentral/
Concluding remarks

• Grave concerns exist over published observational research results, due to study bias, publication bias, and p-hacking

• Large-scale observational studies allow for
  – Empirical evaluation and calibration
  – Unbiased dissemination
  – Providing a more complete evidence base

• LEGEND applies this to real world problems
  – Depression
  – Hypertension