Large-scale evidence generation across a network of databases (LEGEND) for hypertension

Marc A. Suchard, MD, PhD
on behalf of the LEGEND team

Joint Statistical Meetings
3 August 2020
Current knowledge base for hypertension

Head-to-head antihypertensive drug comparisons

- Driven primarily by ALLHAT
  - just 3 individual drugs
- Focus: efficacy $\gg$ safety
- New RCTs too expensive

Trials: 40
$N = 102 - [1148] - 33K$

Can we provide

1. reliable / reproducible – concordant extant w/ RCTs
2. rich – across “all” comparators, outcomes
3. relevant – inform practice evidence?
Observational research estimates in literature

29,982 drug safety estimates from 11,758 papers

What is going wrong?

- Observational bias (confounding, selection, measurement error)
- Publication bias
- $p$-hacking (one study at a time)
- Reproducibility across populations

- 85% have reported $p < 0.05$
- Also note unusual peak along boundary
Large-scale evidence generation across a network of databases (LEGEND)

- Aims to generate reliable evidence on the effects of medical interventions using observational healthcare data
- 10 guiding principles; chief among these:
  - Generate at **large-scale** (completeness, **empirical calibration**)
  - Systematically driven by **best-practices**
  - Disseminate **everything** (open science)

No one person has all the necessary skills
Best-practices: systematic design

**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 THZ

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

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

**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
Observ. study for comparing two initial therapies

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

Medical history lookback time
[PS adjustment]
Follow-up time

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

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

Index: Time zero

Treatment strategies:
- Monotherapy with ACE
- Monotherapy with THZ

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

PS adjustment
Comparison of hypertension treatments

- 39 mono-drugs, 13 mono-classes
- 58 duo-drugs, 32 duo-classes
- 10,278 comparisons

<table>
<thead>
<tr>
<th>Comparison Type</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>
Best-practices: systematic large-scale PS

- >8,000 (regularized) baseline patient characteristics (all dx, rx, tx)
- Address observed (and some unobserved – BP control) confounding (Tian et al, 2019, IJE)
Of course, not all comparisons are valid

- Evaluation of propensity score (PS) distributions and covariate balance
- Here: poor empirical clinical equipoise
Best-practices: 58 expert-crafted outcomes

- Effectiveness (10): acute MI, heart failure, stroke
- Safety (48): known side-effects

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Logical description</th>
<th>Supporting references</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal pain</td>
<td>Abdominal pain condition record of any type; successive records with &gt; 90 day gap are considered independent episodes</td>
<td>4 5 6</td>
</tr>
<tr>
<td>Abnormal weight gain</td>
<td>Abnormal weight gain record of any type; successive records with &gt; 90 day gap are considered independent episodes; note, weight measurements not used</td>
<td>7</td>
</tr>
<tr>
<td>Abnormal weight loss</td>
<td>Abnormal weight loss record of any type; successive records with &gt; 90 day gap are considered independent episodes; note, weight measurements not used</td>
<td>8</td>
</tr>
<tr>
<td>Acute myocardial infarction</td>
<td>Acute myocardial infarction condition record during an inpatient or ER visit; successive records with &gt; 180 day gap are considered independent episodes</td>
<td>9 10 11 12 13 14</td>
</tr>
<tr>
<td>Acute pancreatitis</td>
<td>Acute pancreatitis condition record during an inpatient or ER visit; successive records with &gt; 30 day gap are considered independent episodes</td>
<td>15 16 17 18</td>
</tr>
<tr>
<td>Acute renal failure</td>
<td>A diagnosis of acute renal failure in an inpatient or ER setting; must be at least 30d between inpatient/ER visits to be considered separate episodes</td>
<td>19 20 21 22 23 24 25 26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Theoretical</th>
<th>Observed (n &gt; 2,500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcomes of interest</td>
<td>58</td>
<td>58</td>
</tr>
<tr>
<td>Target-comparator-outcomes</td>
<td>2,843,250 * 58 = 164,908,500</td>
<td>587,020</td>
</tr>
</tbody>
</table>
Calibrate each study (under null)

76 negative outcome controls (not caused by either treatment) help expose and control residual bias. Example: ingrown toenail

- Crude - Uncalibrated: 68% have $p < 0.05$
- PS stratified - Uncalibrated: 16% have $p < 0.05$
- PS stratified - Calibrated: 4% have $p < 0.05$

$p$-value empirical calibration models residual bias as exchangeable and adjusts for a (possibly) non-0 mean. (Schuemie et al, 2018, PNAS)
Network of data sources

- **US insurance databases**
  - IBM MarketScan CCAE
  - IBM MarketScan MDCR
  - IBM MarketScan MDCD
  - Optum Clinformatics
- **Japanese insurance database**: JMDC
- **Korean insurance database**: NHIS-NSC
- **US EHR databases**
  - Optum EHR
  - Columbia University Medical Center
- **German EHR database**: IQVIA DA Germany

Account for population/practice heterogeneity (Madigan et al, 2013, AJE)
Improve generalizability
How does LEGEND perform?

- Best-practices **systematic design, evaluation** and empirical **calibration** return near nominal performance
- Provide a more complete and reliable evidence basis
Unbiased LEGEND dissemination

- Open source protocol and end-to-end executable code
- [http://data.ohdsi.org/LegendBasicViewer](http://data.ohdsi.org/LegendBasicViewer) (all result artifacts for each study)
- [http://data.ohdsi.org/LegendMedCentral](http://data.ohdsi.org/LegendMedCentral) (gimmick)
Head-to-head HTN drug comparisons

- Trials: 40
- $N = 102 - [1148] - 33K$

- Comparisons: 10,278
- $N = 3502 - [212K] - 1.9M$
First-line agents: comparisons from LEGEND

Efficacy outcome: **myocardial infarction**, heart failure, stroke

**RCTs**

<table>
<thead>
<tr>
<th></th>
<th>ACEIs</th>
<th>ARBs</th>
<th>cBBs</th>
<th>dCCBs</th>
<th>THZs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACEIs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARBs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cBBs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dCCBs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THZs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**LEGEND**

<table>
<thead>
<tr>
<th></th>
<th>ACEIs</th>
<th>ARBs</th>
<th>cBBs</th>
<th>dCCBs</th>
<th>THZs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACEIs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARBs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cBBs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dCCBs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THZs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data source: meta-analysis, \( \sim 1 - 2M \) total patients per study

- Beta blockers underperform alternatives
- Unexpected: THZs > ACEs. Reliable?
Statistical analysis. We conduct our cohort study using the open-source OHDSI CohortMethod R package (Schuemie et al. 2018a). The model is trained on a large-scale dataset to estimate the propensity score and controls for observed confounding factors. We use stratification and balance diagnostics to assess the model's performance.

**Large-scale propensity score model controls for observed confounding**

**Cohort stratification / balance:**
- Achieved across all 10,868 baseline characteristics (CCAE)
- Blood pressure (pop. means in mmHg) (Panther)

<table>
<thead>
<tr>
<th></th>
<th>THZs</th>
<th>ACEIs</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>before</td>
<td>145/89</td>
<td>145/87</td>
<td>0.13</td>
</tr>
<tr>
<td>after</td>
<td>145/88</td>
<td>145/87</td>
<td>0.02</td>
</tr>
</tbody>
</table>

No BP measurements used in PS model, but still balanced after stratification.
THZs vs. ACEIs: study outcomes

Calibration returns near nominal HR estimate coverage

- Good diagnostics → comparable cohorts (observed and unobserved); calibration → controls for residual systematic bias
- THZs are more effective than ACEIs in preventing MI
Ideal positioning for COVID-19

Renin-angiotensin system blockers and susceptibility to COVID-19: a multinational open science cohort study

Daniel R Morales, Mitchell M Conover, Seng Chan You, Nicole Pratt, Kristin Kostka, Talita Duarte Salles, Sergio Fernandez Bertolin, Maria Aragon, Scott L DuVall, Kristine Lynch, Thomas Falconer, Kees van Bochove, Cynthia Sung, Michael E. Matheny, Christophe G. Lambert, Fredrik Nyberg, Thamir M AlShammari, Andrew E. Williams, Rae Woong Park, James Weaver, Anthony G. Sena, Martijn J. Schuemie, Peter R. Rijnbeek, Ross D. Williams, Jennifer C.E Lane, Albert Prats Uribe, Lin Zhang, Carlos Areia, Harlan Krumholz, Daniel Prieto Alhambra, Patrick B Ryan, George Hripcsak, Marc A Suchard

doi: https://doi.org/10.1101/2020.06.11.20125849

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.

Abstract

Introduction: Angiotensin converting enzyme inhibitors (ACEs) and angiotensin receptor blockers (ARBs) could influence infection risk of coronavirus disease (COVID-19). Observational studies to date lack pre-specification, transparency, rigorous ascertainment adjustment and international generalizability, with contradictory results. Methods: Using electronic health records from Spain (SIDIAP) and the United States (Columbia University Irving Medical Center and Department of Veterans Affairs), we conducted a systematic cohort

- Over 1.1 million antihypertensive users
- Active comparator, prevalent-users
- Executed within new EHR data partners
  - SIDIAP (universal primary care in Catalonia)
  - US VA
- End-to-end (almost) transparency
Acknowledgments

LEGEND Scientific Group:
- Martijn J. Schuemie
- Patrick B. Ryan
- Seng Chan You
- Nicole Pratt
- David Madigan
- George Hripcsak
- Marc A. Suchard

Clinical Advisory Team:
- RuiJun Chen
- Jon Duke
- Christian Reich
- Harlan Krumholz

Some of these results in Suchard et al. *Lancet*, 2019