Where is the Bigger Gap in RWE?

Innovation to Solve Analytical Challenges VS Implementation of Best Practices with Current Methods

Wei Shen, Eli Lilly & Company
With help from Doug Faries, Xiang Zhang

MBSW 2018
Big Data Quantified

Big Data is made of structured and unstructured information:

- **10% Structured**
  - Structured information is the data in databases and is about 10% of the story.

- **90% Unstructured**
  - Unstructured information is 90% of Big Data and is ‘human information’ like emails, videos, tweets, Facebook posts, call-center conversations, closed circuit TV footage, mobile phone calls, website clicks.

Big Data is only getting bigger:

- 90% of the data in the world today was created within the last two years.
- It will likely reach ~40,000 exabytes or 40 trillion gigabytes by 2020.

Companies are spending big on Big Data:

- In 2015:
  - Financial Services: $6.4B
  - Software Internet: $2.8B
  - Government: $2.8B
  - Comms & Media: $1.2B
  - Energy / Utilities: $800M

Annual Growth to 2020:

- 22%
- 26%
- 22%
- 40%
- 54%
What is Real World Evidence?

Real-World Research Questions + Real-World Design and Analytics + Real-World Data = Real-World Evidence

Primary data sources:
- Patient registries
- Prospective Observational studies
- Pragmatic trials
- Surveys

Secondary data sources:
- EMRs
- Claims data
- Personal health records
- Patients-derived data via smart technologies

Non-interventional research that involves the collection of scientifically valuable, naturalistic, and relevant information for the purpose of answering important research questions.
No Shortage of Guidance

2007 STROBE
• 22 Item Checklist

2009 ISPOR Good Res. Practices
• Design and Reporting (Berger et al); Mitigating Bias (Cox et al); Analytic Methods (Johnson et al)

2010 GRACE
• Dreyer et al (2010); ISPE

2014 PCORI & ISPOR-AMPC-NPC
• Methodology Reports; Flowchart (Berger et al 2014)

2017 Joint ISPOR-ISPE TaskForce

FDA Guidance (2017): Use of RWE .... for Medical Devices
What Does Good Look Like?

Observational Studies – Credibility

Prospective – Retrospective

No → Pre-specification? Yes

No → Adequate Sample? Yes

No → Data: Exposure Outcome Valid? Yes

No → Uncertainty Reported? Yes

No → Methods Reporting Adequate? Yes

No → Sensitivity Analyses Yes

No → Assessment and Control of Confounding? Yes

No → Absolute and Relative Measures Reported Yes

No → Interpretation balanced? Yes

No → Conflict of interest? Yes

No → Dealt with? Yes

ISPOR-AMCP-NPC Good Practice Task Force Report  (Berger et al 2014 Value in Health)
Quality of Current Real-World Evidence (Ryan P, Madigan D. NISS 2016)

What is the quality of the current evidence from observational analyses?

JAMA

Exposure to Oral Bisphosphonates and Risk of Esophageal Cancer

Chris R. Cardwell, PhD
Christian C. Abnet, PhD
Marie M. Cantwell, PhD
Liam J. Murray, MD

Context: Use of oral bisphosphonates has increased over time in the United States, and esophageal cancer is a known adverse effect of these medications. However, the relationship between bisphosphonate use and esophageal cancer has not been well-studied in observational analyses.

Objective: To investigate the association between bisphosphonate use and risk of esophageal cancer in observational studies.

BMJ

Oral bisphosphonates and risk of cancer of oesophagus, stomach, and colorectum: case-control analysis within a UK primary care cohort

Jane Green, clinical epidemiologist,1 Gabriela Zannoli, statistician,1 Gillian Reeves, statistical epidemiologist,1 Joanna Watson, epidemiologist;1 Lesley Wise, manager;2 Pharmacoepidemiology Research Unit,1 Valerie Baril, professor of cancer epidemiology;

ABSTRACT

Objective To examine the observational study of the case-control study to assess the risk of cancer of oesophagus, stomach, and colorectum associated with the use of oral bisphosphonates.

Methods A case-control study of patients with cancer of oesophagus, stomach, and colorectum diagnosed in the UK general practice research database, matched to controls from the same database, was conducted. The exposure of interest was oral bisphosphonate use within the previous 5 years.

Results The risk of cancer of oesophagus, stomach, and colorectum was increased for patients who used oral bisphosphonates compared with those who did not (adjusted odds ratio, 1.2; 95% CI, 1.0 to 1.4). This association was stronger for patients who used bisphosphonates for longer than 5 years (adjusted odds ratio, 1.5; 95% CI, 1.1 to 2.1).

Conclusions The risk of oral bisphosphonate use is associated with an increased risk of cancer of oesophagus, stomach, and colorectum.

August 2010: “Among patients in the UK General Practice Research Database, the use of oral bisphosphonates was not significantly associated with incident esophageal or gastric cancer.”

September 2010: “In this large nested case-control study within a UK cohort [General Practice Research Database], we found a significantly increased risk of oesophageal cancer in people with previous prescriptions for oral bisphosphonates.”
What is the quality of the current evidence from observational analyses?

**BJCP May 2012:** “In this study population, pioglitazone does not appear to be significantly associated with an increased risk of bladder cancer in patients with type 2 diabetes.”

**BMJ May 2012:** “The use of pioglitazone is associated with an increased risk of incident bladder cancer among people with type 2 diabetes.”

**Quality of Current Real-World Evidence**

(Ryan P, Madigan D. NISS 2016)
Challenges with Comparative Real-world Research

• **Lack of randomization** leading to inconsistency, lack of reliability and reproducibility

• **Unknown Operating Characteristics**
  • Do we have a Type 1 error rate of 5%?
  • What do 95% confidence intervals from observational research really mean?
Observational Medical Outcomes Partnership (OMOP)

- Public Private Partnership: FDA, Pharma, Healthcare Providers
- Inform appropriate use of observational data for studying effects (risks and benefits) of medicinal products
- Conduct methodological research to empirically evaluate the performance of various analytical methods on their ability to identify true associations and avoid false findings
  - Extensive simulations to assess the operating characteristics
OMOP Simulations (Ryan et al 2012)

2010-2013 OMOP Research Experiment

- Open-source
- Standards-based

Common Data Model

- 10 data sources
- Claims and EHRs
- 200M+ lives

Positive/Negative Controls

<table>
<thead>
<tr>
<th>Outcome</th>
<th>ACE Inhibitors</th>
<th>Angiotensin B</th>
<th>Antihistamines</th>
<th>Antidepressants</th>
<th>Atorvastatin</th>
<th>Beta Blockers</th>
<th>Bisphosphonates</th>
<th>Benzodiazepines</th>
<th>Triyclic antidepressants</th>
<th>Typical antipsychotics</th>
<th>Warfarin</th>
</tr>
</thead>
</table>
Case-control Estimates for Gl Bleeding Negative Controls
Case-control Estimates for GI Bleeding Negative Controls

- 95% * 65 = 62 of the CIs should cover RR = 1
- We observed 29 of negative controls covered RR=1
- Coverage probability = 45%
- Positive tendency: 74% of estimates have RR > 1
A Call for Improvement

Following best practices is important
• However, despite guidance and much literature the operating characteristics of comparative observational research needs improvement

Need for More Innovation / New Approaches
• Data
• Study Design
• Outcome
• Measure
• Analytical Methods
Analytical Methods for Dealing with Confounding
Types of Bias

- Bias can affect all types of studies though observational studies are particularly vulnerable

- **Confounding**
  - Bias due to a third factor that distorts the association between exposure and outcome.

- **Selection Bias**
  - Bias that results from the selection and retention of the study population.

- **Information Bias**
  - Bias that results from poor measurement of study variables - exposure, outcome, confounders.
Confounding is THE issue

- **With randomization** – standard methods produce estimates of causal treatment effects
- **Without randomization** – standard methods produce only ‘associations’ …. Treatment groups are NOT comparable at baseline thus comparisons are BIASED

#1 Challenge: **Confounding** (Measured and Unmeasured)
Example of Confounding

<table>
<thead>
<tr>
<th>Percent with Event*</th>
<th>RCT (TRITON-TIMI 38)</th>
<th>RWD (Premier Database)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention</td>
<td>Control</td>
</tr>
<tr>
<td>N = 13,457</td>
<td>4.5</td>
<td>3.4</td>
</tr>
<tr>
<td>N = 114,947</td>
<td>2.3</td>
<td>4.1</td>
</tr>
</tbody>
</table>

*RCT: TIMI criteria  
*RWD: ICD-9 criteria

RCT: Prasugrel Label FDA.GOV accessed April 2018
RWD: Ernst et al 2012 QCOR.
Current State of the Union

What should I do about unmeasured confounding?

Just mention it as a limitation in the Discussion Section and move on!
Addressing unmeasured confounding in comparative observational research

Xiang Zhang¹ | Douglas E. Faries¹ | Hu Li¹ | James D. Stamey² | Guido W. Imbens³

Additional information on Unmeasured Confounder ??

None

- Plausibility
  - Negative control
  - Pseudo Trt.
  - Partial identification
  - Empirical Calibration
  - Plausibility Set 1

- Adjusted Analysis
  - IV
  - Difference in differences
  - Regression discontinuity
  - Missing cause
  - Trend-in-trend
  - Perturbation variable

Internal

- Plausibility
  - Plausibility Set 1
  - R & R sensitivity
  - Rosenbaum Sensitivity

- Adjusted Analysis
  - Adjusted Analysis set I
  - Bayesian Twin Regression
  - Multiple imputations
  - Propensity calibration

External

- Plausibility
  - Plausibility Set 1
  - R & R sensitivity
  - Rosenbaum Sensitivity

- Adjusted Analysis
  - Bayesian Twin Regression
  - Propensity calibration
Bayesian Approach: Incorporating Multiple Sources of “External” Information

Comparative effectiveness for treatments for osteoporosis

BMD Information

- Registry
- Public Database
- Literature

BMD Information:
1) strength of association between BMD and treatment selection
2) strength of association between BMD scores and other covariates
3) strength of association between BMD and outcome (fractures)

Truven Claims Analysis

External Data

Adjusted Analysis
Bayesian Twin Regression Models

Fractures = $\beta_0 + \beta_1 \times \text{cohort} + \eta \times \text{covariates} + \lambda \times \text{BMD}$

(2) $\text{BMD} = \gamma_0 + \gamma_1 \times \text{cohort} + \varphi \times \text{covariates}$
Bayesian Twin Regression Analysis

Findings

- Without adjusting for unmeasured BMD, the intervention seems to increase the risk of fracture.
- After incorporating the BMD information, the intervention shows a risk reduction.

* Statistically significant at p<0.05 level.
Limitation: Transportability of Information

Transportability:
Is the BMD Information in these studies relevant to our Truven population?

BMD Information

Registry

Public Database

Literature

Truven Claims Analysis

Bayesian External Data Adjusted Analysis
Analytical Methods for Dealing with Selection Bias
Propensity Score Matching

Propensity Score – the conditional probability that a patient received treatment A given their set of observed baseline covariates X
Basic Assumptions for Causal Inference
(Rubin’s potential outcome framework)

Propensity Score adjustments can provide for estimates of the causal group differences under the following assumptions:

#1 No Unmeasured Confounders
   All confounders are in the dataset and analysis

#2 Sufficient Overlap in Populations
   positivity, no perfect confounding

#3 Correct Statistical Models
Basic Methods for Implementing PS

**Regression**
Simple regression model
Y = Trt + PS

**Stratification**
Group patients with similar PS; Compare cohorts within each PS strata; then average across the strata

**Matching**
Match patients with similar PS, then compare Cohorts of matched pairs

**Inverse Weighting**
Run weighted analysis, weighting each patient by the inverse of their PS
Steps to a Quality Propensity Score Analysis

1. Estimate the Propensity Score
2. Assess Overlap & Balance
3. Estimate the Treatment Effect
4. Sensitivity Analyses

Design without outcome in sight (Rubin 2007)
Use of Propensity Scoring for Comparative Effectiveness is well established ……. Yet Room for Innovation!

Simulations show no Analytical method is best across all Scenarios – No Gold Standard Recommendation (Zagar 2017)

? How do we know which Propensity (or other) method to use in a specific situation?

? How do we know WHEN we have sufficient equipoise for an appropriate analysis?
What if Little Overlap?

Stop? Subset?

Walker (2013): Empirical Equipoise
- “Preference Score” Guidance for when to even continue analysis
Nearest Neighbour (Greedy) Matching

Avg imbalance 0.09

Caliper of .10

Avg imbalance 0.015
## PS Matching Example: Baseline Characteristics (Pawaskar et al 2012)

<table>
<thead>
<tr>
<th></th>
<th>Before Matching</th>
<th>After Matching</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment N=4494</td>
<td>Control N=5424</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>56</td>
<td>61</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Male(%)</td>
<td>39.6</td>
<td>50.2</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>CCI</td>
<td>0.5</td>
<td>0.7</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Diabetes Complic.</td>
<td>19.2</td>
<td>26.3</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Diabetes Comorb.</td>
<td>86.1</td>
<td>67.7</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>A1C</td>
<td>7.7</td>
<td>8.6</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>BMI</td>
<td>38.5</td>
<td>33.7</td>
<td>&lt;.01</td>
</tr>
</tbody>
</table>
### PS Matching Example: Baseline Characteristics (Pawaskar et al 2012)

<table>
<thead>
<tr>
<th></th>
<th>Before Matching</th>
<th>After Matching</th>
<th>P-value</th>
<th>Before Matching</th>
<th>After Matching</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment N=4494</td>
<td>Control N=5424</td>
<td>P-value</td>
<td>Treatment N=2683</td>
<td>Control N=2683</td>
<td>P-value</td>
</tr>
<tr>
<td>Age</td>
<td>56</td>
<td>61</td>
<td>&lt;.01</td>
<td>58</td>
<td>58</td>
<td>.28</td>
</tr>
<tr>
<td>Male(%)</td>
<td>39.6</td>
<td>50.2</td>
<td>&lt;.01</td>
<td>45.4</td>
<td>47.1</td>
<td>.20</td>
</tr>
<tr>
<td>CCI</td>
<td>0.5</td>
<td>0.7</td>
<td>&lt;.01</td>
<td>0.6</td>
<td>0.6</td>
<td>.69</td>
</tr>
<tr>
<td>Diabetes Complic.</td>
<td>19.2</td>
<td>26.3</td>
<td>&lt;.01</td>
<td>21.8</td>
<td>22.0</td>
<td>.84</td>
</tr>
<tr>
<td>Diabetes Comorb.</td>
<td>86.1</td>
<td>67.7</td>
<td>&lt;.01</td>
<td>80.4</td>
<td>80.8</td>
<td>.74</td>
</tr>
<tr>
<td>A1C</td>
<td>7.7</td>
<td>8.6</td>
<td>&lt;.01</td>
<td>8.1</td>
<td>8.1</td>
<td>.63</td>
</tr>
<tr>
<td>BMI</td>
<td>38.5</td>
<td>33.7</td>
<td>&lt;.01</td>
<td>36.8</td>
<td>36.6</td>
<td>.56</td>
</tr>
</tbody>
</table>
Use of RWE growing

Progress on guidance, establishing and implementing best practices are important

Innovation and research to resolve current challenges & demonstrate improved operating characteristics is essential