Social Determinants and EHR Data: Analytic Decision Support

Harold P. Lehmann MD PhD
The PaTH Clinical Data Research Network
Patrick Ryan, Observational Health Data Sciences and Informatics (OHDSI) Overview, 5/14/14

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.”

 Courtesy Kelly Gleason
Observational Health Data Sciences and Informatics (OHDSI) Overview, 5/14/14

April 2012: “Patients taking oral fluoroquinolones were at a higher risk of developing a retinal detachment”

Dec 2013: “Oral fluoroquinolone use was not associated with increased risk of retinal detachment”

Courtesy Kelly Gleason
How do I convince hard-boiled researchers that our results are as trustworthy and believable as the best epidemiological data?

Dan Ford
Where’s the Population?

What’s the “diagnosis”?
The case atrial fibrillation

• Do all the fields with the same name mean the same thing?

**Diagram:**
- **Encounter:** 15,774
- **Billing:** 18,731
- **Problem List:** 25,608
- Distinct values:
  - Encounter: 850
  - Billing: 4,500
  - Problem List: 11,054
  - Overlapping values:
    - Encounter & Billing: 14,456
    - Encounter & Problem List: 2,659
    - Billing & Problem List: 1,986
  - Common values: 33,314
Some Potential Biases

- Diagnostic/Treatment Access Bias
- Healthcare Access Bias
- Referral Filter Bias
- Spectrum Bias
- Berkson’s Bias
- Length Bias
- Diagnostic Suspicion Bias
- (Survivor) Treatment Bias
- Lead time/Protopathic Bias
- Inclusion/Exclusion Bias
- Spectrum Bias
- Inclusion/Exclusion Bias
- Under-reporting/Recall Bias
- Semantic Uncertainty
- Non-Response Bias
- Sick-Quitter Bias
- Gaps in Data
- Recorded Outcome
- Temporal Ambiguity
- Patient-Reported Outcome
- Observed Outcome
- Observable Outcome
- Computable Cohort <Exposure>
- Centripetal Bias
- Lead time/Protopathic Bias
- Competing Risks
- Spectrum Bias
- Ascertertainment/Misclassification/Detection Bias
- General Population
- EHR Population
- Healthcare Population
- General Population

DHSI
Division of Health Sciences Informatics
Amateur Analysts

• Too many analysts to train them all at the level we want
  ➢ MACRA, eCQM, Pop Health, PMI, ...

• Analyses are the most complicated

• No funds for proper statistical analysis

• Statistical-analytic decision support is needed

• We need to convert methodological knowledge into computer-readable form
“According to the McKinsey report, the United States will need an additional 140,000 to 190,000 data science experts with “deep analytical skills,” plus 1.5 million managers capable of using data analytics in decision making.
Decision Support Cycle

- Expert knowledge
- Scientific literature
- Enterprise data warehouse

Knowledge base

Patient data
MD info

deployed

Action

Knowledge engineering
Knowledge use
Decision Support Cycle

- Expert knowledge
- Scientific literature
- Enterprise data warehouse

Knowledge base

- Data set
- Analyst’s Knowledge

Knowledge engineering

Knowledge use
“By 1995 or so, the largest single driving force in guiding general work on data analysis and statistics [will be] to understand and improve data-analytic expert systems…”

John Tukey, 1986
Early History

• 1983: Nedler: Front-end system (for GLIM)
• 1984: Gale, Pregiborn: REX: Advise on linear regression
• 1985: Hahn defines levels of intelligence: simple computerized answering → automated statistical consulting
• 1988: Duijsens: PRINCE helps naïve users formulate analysis options
• 1988: Oldford & Peters: DINDE: graphical environment tracks steps
• 1989: Chowdury: MAXITAB for inexperienced users for data analysis and interpretation
• 1994: Silvers et al.: PROPHET: Beyond Anova
• Silvers, 1994
<table>
<thead>
<tr>
<th>Desiderata</th>
<th>Development Work to Be Done</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Discrimination</td>
<td>• Measures that take clinical thresholds into account(^{70,71})</td>
</tr>
<tr>
<td></td>
<td>• Elicitation and articulation of those thresholds</td>
</tr>
<tr>
<td></td>
<td>• Methods for recalculating local discrimination</td>
</tr>
<tr>
<td>2. Local Recalibration</td>
<td>• Application of calibration based on thresholds(^{17})</td>
</tr>
<tr>
<td>3. Thresholds &amp; Local Preferences</td>
<td>• Elicitation, articulation of preferences</td>
</tr>
<tr>
<td></td>
<td>• Local calculation of thresholds</td>
</tr>
<tr>
<td>4. Explanation</td>
<td>• Deployment</td>
</tr>
<tr>
<td>5. Monitoring</td>
<td>• Choose variables based on value of information(^{72})</td>
</tr>
<tr>
<td>6. Debiasing</td>
<td>• Creation and curation of debiasing models</td>
</tr>
<tr>
<td></td>
<td>• Application of debiasing models</td>
</tr>
<tr>
<td>7. Generalizability</td>
<td>• Calculation of distance(^{62})</td>
</tr>
<tr>
<td></td>
<td>• Adding to the Knowledge Artifact the meta data required to choose the calculation</td>
</tr>
<tr>
<td>8. Semantic Uncertainty</td>
<td>• Derivation of the epistemic confidence interval</td>
</tr>
<tr>
<td>9. Findable</td>
<td>• Articulation of the full ontology required to index a Knowledge Artifact at all its multiple levels</td>
</tr>
<tr>
<td></td>
<td>• Tagging KO with that ontology</td>
</tr>
<tr>
<td>10. Other Commandments as necessary and proper</td>
<td>• Continuous monitoring and improvement of these desiderata</td>
</tr>
</tbody>
</table>
Ontology for Biases: Extensions to OCRe

To Do

- Methodology for the analysts
- Knowledge tools to store the knowledge
- Knowledge tools to apply the knowledge
- Combine JH/PaTH/Israeli expertise