A Data-Driven Education: Learning Medicine by Understanding Bias in Large, De-identified EHR Databases

VISHAL N. PATEL, MD PHD
INSTRUCTOR & RESEARCH FELLOW
VXP43@CASE.EDU
EHR Databases

SNOMED-CT

Fast Search
Explore: Universe

Browse
Demographics | Diagnoses | Findings | Observe | Drugs | Procedures | Vitals

- Disorder of trunk
- Disorder of body cavity
- Disorder of soft tissue
- Disorder of cardiovascular system
- Soft tissue lesion
- Disorder of abdominal segment of trunk
- Disorder of abdomen
- Metabolic disease
- Disorder of thoracic segment of trunk
- Disorder of thorax

Count | Proportion
---|---
208,420 | 85%
199,710 | 82%
191,370 | 78%
185,990 | 76%
181,620 | 74%
176,800 | 73%
170,710 | 70%
162,050 | 66%
161,760 | 66%
160,460 | 66%

Cohort
Population Size: 243,790
Universe

Cohort: Unsaved

- Concept
- Diagnosis: Rheumatoid arthritis
  Size: 243,790

- AND

- Any of these . . . (or)

Drag items here

Save | Clear | Convert to Power Search

Export To Excel file
Bias Detection as Learning

Covariates:
- Sign/symptom
- Cause
- Effect
- Confounder
- Selection bias

Data Mining:
- Odds/risk ratios, p-values
- Stratified risk ratios
- Network models

Browse

<table>
<thead>
<tr>
<th>Description</th>
<th>Count</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disorder of bunk</td>
<td>208,420</td>
<td>56%</td>
</tr>
<tr>
<td>Disorder of body cavity</td>
<td>169,710</td>
<td>52%</td>
</tr>
<tr>
<td>Disorder of soft tissue</td>
<td>161,370</td>
<td>53%</td>
</tr>
<tr>
<td>Disorder of cardiovascular system</td>
<td>155,300</td>
<td>75%</td>
</tr>
<tr>
<td>Soft tissue lesion</td>
<td>161,820</td>
<td>74%</td>
</tr>
<tr>
<td>Disorder of abdominal segment of trunk</td>
<td>176,800</td>
<td>73%</td>
</tr>
<tr>
<td>Disorder of abdomen</td>
<td>170,710</td>
<td>70%</td>
</tr>
<tr>
<td>Metabolic disease</td>
<td>162,050</td>
<td>88%</td>
</tr>
<tr>
<td>Disorder of thoracic segment of trunk</td>
<td>161,760</td>
<td>66%</td>
</tr>
<tr>
<td>Disorder of thorax</td>
<td>160,460</td>
<td>56%</td>
</tr>
</tbody>
</table>

Odds/risk ratios, p-values
Stratified risk ratios
Network models
Medical Education

- Evidence-based Medicine \(\rightarrow\) Comparative Effectiveness Research
  1. Built on EHR data

- Learning (Ertmer & Newby 1993)
  1. Process of creating meaning
  2. Building interpretations
  3. Self-directed