PheKnow-Cloud: A Tool for Evaluating High-Throughput Phenotype Candidates using Online Medical Literature

Jette Henderson¹, Ryan Bridges², Joyce C. Ho³, Byron C. Wallace⁴, Joydeep Ghosh¹

¹UT Austin, ²Epic Systems, ³Emory University, ⁴Northeastern University

AMIA Joint Summits on Translational Sciences
Disclosure

• Neither my collaborators nor I have no relationships with commercial interests or conflicts of interests
Background: EHR-Based Phenotyping

Manual Phenotype Extraction
- Laborious
- Time-consuming
- Requires domain expertise
Motivation & Background

Automatic Phenotype Generation Process

Raw Electronic Healthcare Record Data

Verification by Panel of Experts

Candidate Phenotypes

Verified Phenotypes

Phenotype 1
- Coronary Artery Disease
- Other Chronic Ischemic Heart Disease
- Other Endocrine Disorders
- Diabetes Unspecified
- Hypertension
- Angina Pectoris
- Myocardial Infarction
- Surgical Procedures on the Musculoskeletal System

Phenotype 2
- Major Symptoms, Abnormalities
- Heart Rhythms
- Other Symptoms and Signs
- Hypertension
- Other Endocrine Disorders
- Other Symptoms and Signs

Phenotype n
- Iron Deficiency and Other Unspecified Anemias
- Other Gastrointestinal Disorders
- Osteoarthritis of Hip or Knee
- Hypertension
- Surgical Procedures on the Musculoskeletal System

Phenotype n
- Major Symptoms, Abnormalities
- Heart Rhythms
- Other Symptoms and Signs
- Hypertension
- Other Endocrine Disorders
- Other Symptoms and Signs

Phenotype n
- Iron Deficiency and Other Unspecified Anemias
- Other Gastrointestinal Disorders
- Osteoarthritis of Hip or Knee
- Hypertension
- Surgical Procedures on the Musculoskeletal System
Goal

Automatic Phenotype Generation Process

Candidate Phenotypes

Verified Phenotypes

Raw Electronic Healthcare Record Data
Raw Electronic Healthcare Record Data → Automatic Phenotype Generation Process → PhenKnow-Cloud → Evidence and Lift Analysis → Phenotype Evidence Results
PheKnow-Cloud Result Interface
Phenotype Verification Process using Pubmed

Phenotype Representation Feature Extraction

Co-occurrence counts on PubMed

Phenotype significance calculation
Feature Extraction

Phenotype 33
- cardiac dysrhythmias
- heart failure
- atrial fibrillation and flutter
- unspecified chest pain
- cardiomyopathy
- calcium channel blocking agents
- loop diuretics
- antianginal agents

MeSH
- "heart failure"
- ICD-10
- SNOMED-CT

"left heart failure", "heart failure", "congestive heart disease", "myocardial failure", ...

SNOMED-CT
- Ranking synonyms and related concepts
- "heart failure", "left heart failure", "myocardial failure", "congestive heart disease", ...

MeSH
- "antianginal agents"
- ICD-10
- SNOMED-CT

"amiodarone 300mg 10ml solution", "form amiodarone", "antianginal agents", "amiodarone hydrochloride", ...

SNOMED-CT
- Ranking synonyms and related concepts
- "antianginal agents", "amiodarone hydrochloride", "form amiodarone", "amiodarone 300mg 10ml solution"

Phenotype

Generation of Candidates for "Synonym Set"

Ranked n-grams for each Phenotypic item
Co-Occurrence Calculation Process

Given terms $A$, $B$, $C$

\[
lift(A, B, C) = \frac{P(A \cap B \cap C)}{P(A) \cdot P(B) \cdot P(C)}
\]
# Significance Determination—Aggregating Lifts within Phenotypes

## Phenotype
- Rheumatoid arthritis and other inflammatory polyarthropathies
- Other and unspecified disorders of joint
- Osteoarthritis and allied disorders
- Osteoporosis
- Hypertension
- Other and unspecified disorders of back
- Miscellaneous analgesics
- Antirheumatics
- Vitamins
- Cox-2 inhibitors
- Glucocorticoids
- Proton pump inhibitors
- Nutraceutical products

## Phenotypic Term
- Hypertension
- Osteoporosis
- Miscellaneous analgesics

## Most Relevant Synonyms
- "regarding hypertension", "hypertensive disorder systemic"
- "ischemia due hypertension", etc.
- "osteoarthritis and allied disorders", "osteoarthritis"
- "osteoarthritis", "rheumatoid arthritis and other inflammatory polyarthropathies"
- "painful periods", "abdominal pain finding", "pain observable entity", etc.

## Co-occurrence and Lift Calculation

<table>
<thead>
<tr>
<th>Co-occurrence set (represented by original phenotypic terms)</th>
<th>Cardinality</th>
<th>Standard Deviations Above Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>('hypertension', 'osteoporosis')</td>
<td>2</td>
<td>0.1169</td>
</tr>
<tr>
<td>('cox-2 inhibitors', 'proton pump inhibitors', 'vitamins')</td>
<td>3</td>
<td>0.0907</td>
</tr>
<tr>
<td>('cox-2 inhibitors', 'osteoporosis')</td>
<td>2</td>
<td>-0.0071</td>
</tr>
<tr>
<td>'osteoarthritis and allied disorders', 'osteoarthritis'</td>
<td>2</td>
<td>0.0053</td>
</tr>
<tr>
<td>('osteoarthritis', 'proton pump inhibitors')</td>
<td>2</td>
<td>0.0266</td>
</tr>
<tr>
<td>('osteoporosis', 'rheumatoid arthritis and other inflammatory polyarthropathies')</td>
<td>2</td>
<td>-0.0018</td>
</tr>
</tbody>
</table>

### Average Standard Deviations Above Median:

.0367
Experimental Set-up

**Phenotype Data**
- Random and curated phenotypes
- 80 annotated phenotypes generated by two different automatic phenotype generation algorithms
  - 14% -- clinically meaningful
  - 78% -- possibly significant
  - 8% not clinically meaningful

**PubMed Data**
- 25% of PubMed Open Access Subset

**Method**
1. Calculate lift
2. Determine “optimal” threshold that separates “significant” and “not significant” phenotypes
Process Tuning—Phenotypic Item Synonym Set Size
Results—Randomly Generated vs Curated Phenotypes

Classification Results

- 100% true negative classification
- 80% true positive classification
- F1 score of 0.89
Results—Automatically Generated Phenotypes

Classification Results
• Threshold = 0.028
• F1 score of 0.87
### Candidate Phenotypes

- Diuretics of fluid, electrolytes, and acid-base balance
- Other medications that influence sodium and osmotic solutes
- Calcium channel blocking agents
- Secondary hypertension
- Selective immune suppressants
- Angiotensin converting enzyme inhibitors
- Hypertension

### Table of Evidence

<table>
<thead>
<tr>
<th>Index</th>
<th>Paper</th>
<th>Standard Deviation above Median</th>
<th>L/R</th>
<th>Co-occurrence Tuples</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Title: Unraveling needs in lupus nephritis solutions through evidence-based, personalized medicine</td>
<td>0.001</td>
<td>calcium channel blocking agents, selective immune suppressants</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Title: Assessment of the Effects of Low-Level Laser Therapy on the Thrombosis Mediation of Platelets in Patients with Autonomic Hypertension by the Doppler Ultrasound Method</td>
<td>0.0196</td>
<td>antiplatelet agents, selective immune suppressants</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Title: Fluid and Electrolyte Fluids in Critically II Patients</td>
<td>0.0004</td>
<td>Changes of fluid, electrolytes, and acid-base balance, hypertension, secondary hypertension</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Title: The Effects of Deceleration on Blood Pressure in Pediatric Patients with Juvenile Insulin-Dependent Diabetes</td>
<td>0.001</td>
<td>Hypertension, electrolytes, secondary hypertension</td>
<td></td>
</tr>
</tbody>
</table>

**This candidate phenotype has an overall standard deviation below the median: lift of 0.0004**
Genetic/metabolic risk stratification, combination of low-dose immunosuppressants with anti-inflammatory drugs, favor specific drugs over unselective immunosuppressants.
Fluid and Electrolyte Disturbances in Critically Ill Patients

Jay Wook Lee, M.D.

(Disorders of fluid, electrolyte, and acid-base balance, hypertension, secondary hypertension)
Future Work

• Incorporate “gold standard” phenotypes from PheKB and other sources
• Scale to whole PubMed Open Access Subset
• Speed up co-occurrence analysis
• Refine and automate phenotype classification process
References


References


