A Method for Probing Disease Relatedness Using Common Eligibility Features in Clinical Trials

Mary R Boland\(^+\), MA, Riccardo Miotto\(^+\), PhD, Chunhua Weng, PhD\(*\)
Department of Biomedical Informatics
Columbia University, New York City
(+: equal contribution first author; *: senior author)
This paper makes two contributions:

- Applies a method using common eligibility features mined from many studies to profile target populations for research on any disease topic

- Establishes clinical trial eligibility criteria as a potential new data source for investigating patterns in subject selection for disease research and for inferring disease relatedness
Example Clinical Trial Eligibility Criteria

• Known HbA1c (patient report or available records at time of enrollment) < 7.5% within prior 6 months
• Advanced visual acuity loss in both eyes which prohibits ability to read study materials (tested as needed with reading test using materials in appropriate size script)
• Subjects with impaired fasting glucose: blood glucose ≥ 1.10 g/l and < 1.26 g/L
• History of diabetes-related medical complications
Research Hypotheses

H1: all clinical trials on any selected disease share a set of common eligibility features (CEFs)

H2: clinical trials on related diseases share similar CEFs
Common Eligibility Features (CEFs) Extraction

Eligibility Criteria Processing
- POS tagger
- n-gram representation
- n-gram filtering
- UMLS lexicon

Feature Selection
- frequent n-grams (i.e. 5%)
- group nested features
- discard non-discriminative frequent features
- group similar features

Feature Post-Processing


ClinicalTrials.gov

Disease-specific Common Eligibility Features
- type II diabetes
- active malignancy
- pregnant
- brain metastasis
...
# Example CEFs for Breast Cancer

<table>
<thead>
<tr>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>platelets adverse event</td>
<td>finding of platelet count</td>
</tr>
<tr>
<td>ecog performance status</td>
<td>ecog performance status</td>
</tr>
<tr>
<td>creatinine</td>
<td>bilirubin</td>
</tr>
<tr>
<td>aspartate transaminase</td>
<td>neoplasm metastasis</td>
</tr>
<tr>
<td>breast carcinoma</td>
<td>prior chemotherapy</td>
</tr>
<tr>
<td>estrogen receptors</td>
<td>creatinine clearance</td>
</tr>
<tr>
<td>pregnancy test negative</td>
<td>electrocorticogram</td>
</tr>
<tr>
<td>pharmaceutical adjuvants</td>
<td>malignant neoplasm of skin</td>
</tr>
<tr>
<td>alanine transaminase</td>
<td>erbb-2 receptor</td>
</tr>
<tr>
<td>operative surgical procedures</td>
<td></td>
</tr>
<tr>
<td>phosphoric monoester hydrolases</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>Disorders</td>
</tr>
<tr>
<td></td>
<td>Laboratory and Tests</td>
</tr>
<tr>
<td></td>
<td>Procedures</td>
</tr>
<tr>
<td></td>
<td>Chemical and Drugs</td>
</tr>
<tr>
<td></td>
<td>Others</td>
</tr>
</tbody>
</table>
Data

- 3 common disease categories and 7 diseases in each category
- Disease trials from ClinicalTrials.gov

Reference: CDC and NINDS

<table>
<thead>
<tr>
<th>Mental Disease</th>
<th>Cancer</th>
<th>Chronic Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epilepsy</td>
<td>Breast cancer</td>
<td>Diabetes</td>
</tr>
<tr>
<td>Huntington’s disease</td>
<td>Prostate cancer</td>
<td>Chronic kidney disease</td>
</tr>
<tr>
<td>Parkinson’s disease</td>
<td>Lung cancer</td>
<td>Arthritis</td>
</tr>
<tr>
<td>Stroke</td>
<td>Colorectal cancer</td>
<td>Obesity</td>
</tr>
<tr>
<td>Amyloid Lateral Sclerosis (ALS)</td>
<td>Liver cancer</td>
<td>atherosclerosis</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>Pancreatic cancer</td>
<td>hypertension</td>
</tr>
<tr>
<td>Depression</td>
<td>Skin cancer</td>
<td>Crohn’s disease</td>
</tr>
</tbody>
</table>
ClinicalTrials.gov as of 12/2012: 13,905 cancer trials, 11,845 chronic disease trials and 5,027 mental disorder trials
CEFs Sharing by Disease Category
Cancers

Chronic diseases

Mental disorders

CEF associated with at least one disease
Crohn’s Disease
Artherosclerosis
Obesity
Chronic Kidney Disease (CKD)
Hypertension
Diabetes
Arthritis
Crohn’s Disease
Arthritis
Chronic Kidney Disease (CKD)
Hypertension
Diabetes
Amyloid Lateral Sclerosis (ALS)
Stroke
Epilepsy
Depression
Parkinson's Disease
Schizophrenia
Huntington's Disease
Disease-specific CEFs shared among mental disorder trials
Findings

✓ H1: clinical trials on a selected disease share CEFs
✓ H2: clinical trials on related diseases share similar CEFs

Also:

◦ The cluster of cancer trials are more cohesive than the cluster of chronic disease trials
◦ Some diseases are closer than other diseases within the same cluster
Conclusions

- CEFs are effective ways to profile disease research patients

- We establish clinical research eligibility criteria as a potential new data source for probing disease relatedness, as opposed to using literature or clinical data

- The degree of CEFs sharing in research recruitment may indicate the degree of relatedness of diseases

- The lack of CEFs sharing in related diseases may uncover other aspects of the research enterprise
Next steps…

Expand analysis to all 150,000 trials on CT.gov

Explore valuable unknown disease relatedness using this network approach

Understand discrepancies in using CEFs for patient recruitment across different disease categories and study if they are due to knowledge sharing and reuse barriers, disease heterogeneity, or research biases.
Acknowledgments

- **NLM R01LM009886** (Bridging the semantic gap between research eligibility criteria and clinical data, PI: Weng)

Thank you!

Chunhua Weng, PhD, cw2384@columbia.edu
Efficiency of CEFs

![Graph showing the efficiency of CEFs with different semantic levels.](image-url)
Stability of CEFs

![Graph showing the stability of CEFs with different R values. The x-axis represents the number of trials, ranging from 5000 to 65000, and the y-axis represents the percentage of tags in common. The graph includes lines for R = 0.5, R = 1.0, R = 3.0, and R = 5.0.]