Presenter Financial Disclosure

I do not have any relationships to report within the last 24 months with ACCME defined ineligible companies.
Unlabeled/Investigational Uses

I will not be discussing unlabeled/investigational uses of medical devices or pharmaceuticals during this presentation.
Age-dependent topic modelling of comorbidities in UK Biobank identifies disease subtypes with differential genetic risk

ASHG 2022
Xilin Jiang
Harvard University/University of Cambridge
• Background: EHR system collects longitudinal patient medical history

• Methods

  • Age-dependent topic modelling of disease records infer latent comorbidities that represent the full patient diagnosis history

• Results

  • Identify distinct comorbidity profiles that are associated with different age-at-onsets.

  • Comorbidity profiles identify genetic heterogeneity between disease subtypes.

• Discussion: the role of individual-level comorbidity for identifying shared and distinct genetic pathways
Longitudinal diagnosis history data are available in large data sets.

EHR data provides individual-level age-at-onset information.

**Individual ID:** 4427213

<table>
<thead>
<tr>
<th>Diagnosis age</th>
<th>ICD-10 diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>K449</td>
</tr>
<tr>
<td>42</td>
<td>I251</td>
</tr>
<tr>
<td>42</td>
<td>I251</td>
</tr>
<tr>
<td>47</td>
<td>R55</td>
</tr>
<tr>
<td>48</td>
<td>T782</td>
</tr>
<tr>
<td>51</td>
<td>S619</td>
</tr>
<tr>
<td>52</td>
<td>N390</td>
</tr>
<tr>
<td>52</td>
<td>Z092</td>
</tr>
</tbody>
</table>

Latent comorbidity weights

---

5
Longitudinal diagnosis history data are available in large data sets.

EHR data provides individual-level age-at-onset information.

**Individual ID:** 4427213

<table>
<thead>
<tr>
<th>Diagnosis age</th>
<th>ICD-10 diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>K449</td>
</tr>
<tr>
<td>42</td>
<td>I251</td>
</tr>
<tr>
<td>42</td>
<td>I251</td>
</tr>
<tr>
<td>47</td>
<td>R55</td>
</tr>
<tr>
<td>48</td>
<td>T782</td>
</tr>
<tr>
<td>51</td>
<td>S619</td>
</tr>
<tr>
<td>52</td>
<td>N390</td>
</tr>
<tr>
<td>52</td>
<td>Z092</td>
</tr>
</tbody>
</table>

**Age-at-onset analysis**
• Background: EHR system collects longitudinal patient medical history

• Methods
  
  • **Age-dependent topic modelling of disease records infer latent comorbidities that represent the full patient diagnosis history**

• Results
  
  • Identify distinct comorbidity profiles that are associated with different age-at-onsets.
  
  • Comorbidity profiles identify genetic heterogeneity between disease subtypes.

• Discussion: the role of individual-level comorbidity for identifying shared and distinct genetic pathways
Infer latent comorbidities to represent patient level diagnosis history
Infer latent comorbidities to represent patient level diagnosis history
Infer latent comorbidities to represent patient level diagnosis history
Infer latent comorbidities to represent patient level diagnosis history.
• Background: EHR system collects longitudinal patient medical history

• Methods

  • Age-dependent topic modelling of disease records infer latent comorbidities that represent the full patient diagnosis history

• Results

  • Identify distinct comorbidity profiles that are associated with different age-at-onsets.

  • Comorbidity profiles identify genetic heterogeneity between disease subtypes.

• Discussion: the role of individual-level comorbidity for identifying shared and distinct genetic pathways
Simulation analysis shows the model could capture age dependent comorbidity.
Simulation analysis shows the model could capture age dependent comorbidity.
• Background: EHR system collects longitudinal patient medical history

• Methods
  • Age-dependent topic modelling of disease records infer latent comorbidities that represent the full patient diagnosis history

• Results
  • Identify distinct comorbidity profiles that are associated with different age-at-onsets.
  • Comorbidity profiles identify genetic heterogeneity between disease subtypes.

• Discussion: the role of individual-level comorbidity for identifying shared and distinct genetic pathways
Application to UK Biobank identifies 10 comorbidity trajectories

1.7 million records across 348 diseases in the UK Biobank

Individual ID: 4427213

<table>
<thead>
<tr>
<th>Diagnosis age</th>
<th>K449 → I251 → I251 → R55 → T782 → S619 → N390 → Z092</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICD-10 diagnosis</td>
<td>35 → 42 → 42 → 47 → 48 → 51 → 52 → 52</td>
</tr>
</tbody>
</table>
Early onset and late reporting of the same disease are associated with distinct comorbidities

CER: circulatory system, endocrine/metabolic, and respiratory diseases

CVD: circulatory system diseases

Topic loading examples inferred from the UK Biobank.
### Age group
- **Y** Young group: < 60
- **O** Old group: ≥ 60

### Posterior topic weights

#### NRI: neoplasm, respiratory, infection
- Colon cancer
- Malignant neoplasm of rectum, rectosigmoid junction, and anus
- Breast cancer [female]
- Malignant neoplasm of female breast
- Cervical intraepithelial neoplasia [CIN] [Cervical dysplasia]
- Lipoma of skin and subcutaneous tissue
- Hypothyroidism NOS
- Type 2 diabetes
- Hypercholesterolemia
- Obesity
- Other anemias
- Major depressive disorder
- Anxiety disorder
- Migraine
- Other peripheral nerve disorders
- Essential hypertension
- Peripheral vascular disease, unspecified
- Phlebitis and thrombophlebitis of lower extremities
- Varicose veins of lower extremity
- Pneumonia
- Pneumococcal pneumonia
- Asthma
- Bronchiectasis
- Postinflammatory pulmonary fibrosis
- Respiratory failure
- Other diseases of respiratory system, NEC

#### LGI: lower gastrointestinal

#### FGND: Female genitourinary, neoplasm, digestive
• Background: EHR system collects longitudinal patient medical history

• Methods

  • Age-dependent topic modelling of disease records infer latent comorbidities that represent the full patient diagnosis history

• Results

  • Identify distinct comorbidity profiles that are associated with different age-at-onsets.

  • Comorbidity profiles identify genetic heterogeneity between disease subtypes.

• Discussion: the role of individual-level comorbidity for identifying shared and distinct genetic pathways
• Comorbidity profiles are inferred using only diagnosis data, not genetic data.

• Genetic data could be utilised to verify the disease subtypes.
PRS are associated with comorbidity weights
Disease PRS are associated with comorbidity weights across cases

CVD: Cardiovascular disease
MGND: Male genitourinary
CER: Type 2 diabetes topic (endocrine)
FGND: Female genitourinary

- Hypercholesterolemia
- Type 2 diabetes
- Asthma
- Essential hypertension

Excess PRS in cases

* FDR < 0.1
** FDR < 0.01
Disease PRS stratification power varies across disease subtypes

CVD: Cardiovascular disease
MGND: Male genitourinary

CVD  MGND  CER  FGND

Disease: Type 2 diabetes

Disease: Hypercholesterolemia
• Background: EHR system collects longitudinal patient medical history

• Methods

  • Age-dependent topic modelling of disease records infer latent comorbidities that represent the full patient diagnosis history

• Results

  • Identify distinct comorbidity profiles that are associated with different age-at-onsets.

  • Comorbidity profiles identify genetic heterogeneity between disease subtypes.

• Discussion: the role of individual-level comorbidity for identifying shared and distinct genetic pathways
Genetic heterogeneity across subtypes using $F_{st}$ and Genetic correlations

P-values for $F_{st}$ analysis of diseases stratified by topic liabilities. 29 diseases have excessive $F_{st}$ across subtypes identified by Comorbidity.
• Comorbidity profiles could be considered as a context for patient health status.
Genetic risk could be modulated by comorbidity context
SNPs have varying effect across quartiles of comorbidity weights

rs1063192 x CVD for Type 2 diabetes
chr9: 9p21.3 (nearest gene:CDKN2B)
P = 3.534e-07 for interaction

We identified 43 SNP x topic interactions at FDR<0.1
• Background: EHR system collects longitudinal patient medical history

• Methods

• Age-dependent topic modelling of disease records infer latent comorbidities that represent the full patient diagnosis history

• Results

• Identify distinct comorbidity profiles that are associated with different age-at-onsets.

• Comorbidity profiles identify genetic heterogeneity between disease subtypes.

• Discussion: the role of individual-level comorbidity for identifying shared and distinct genetic pathways
Individual-level comorbidity for identifying shared and distinct genetic pathways

1. Individuals with same disease code have distinct comorbidities.

2. The comorbidity heterogeneity among the cases imply the distinct disease subtypes.

3. The individual disease context captured by comorbidity could modulate the genetic risk effect sizes.
MedRxiv: Age-dependent topic modelling of comorbidities in UK Biobank identifies disease subtypes with differential genetic risk

Acknowledgement:

Gil McVean  Alkes Price  Chris Holmes  Martin Zhang  Yidong Zhang

& Members of Alkes group

Funding:

Wellcome Trust studentship
Rhodes Scholarship

Thank you!

xilinjiang@hsph.harvard.edu