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

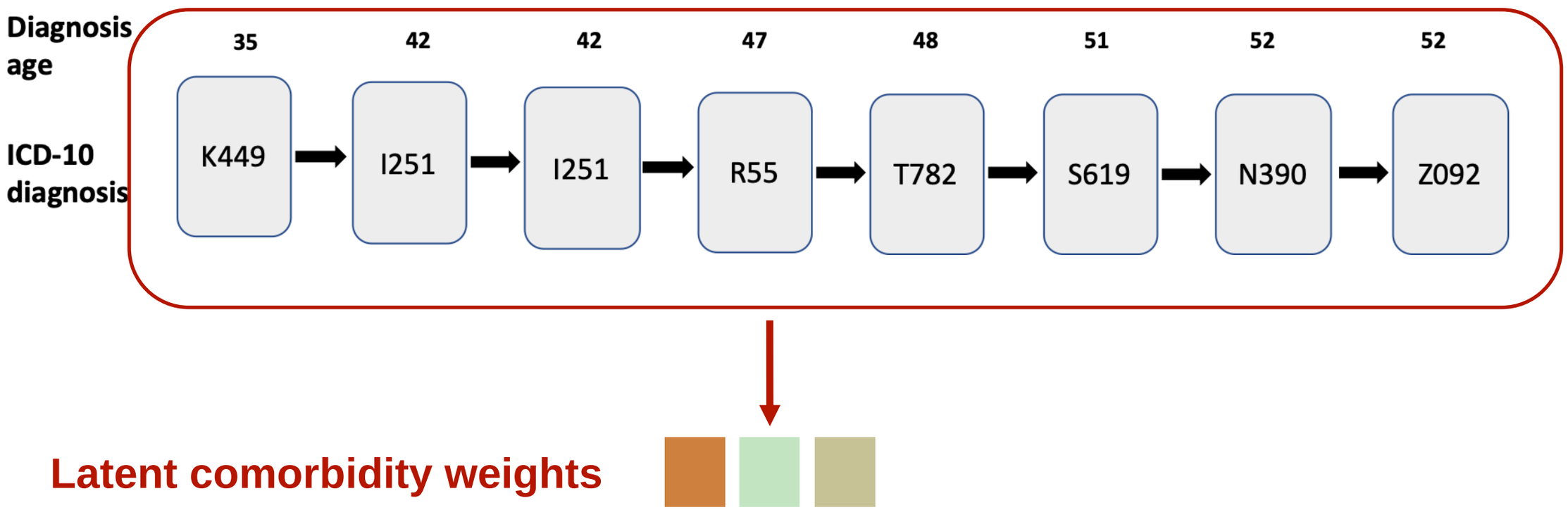
ASHG 2022

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

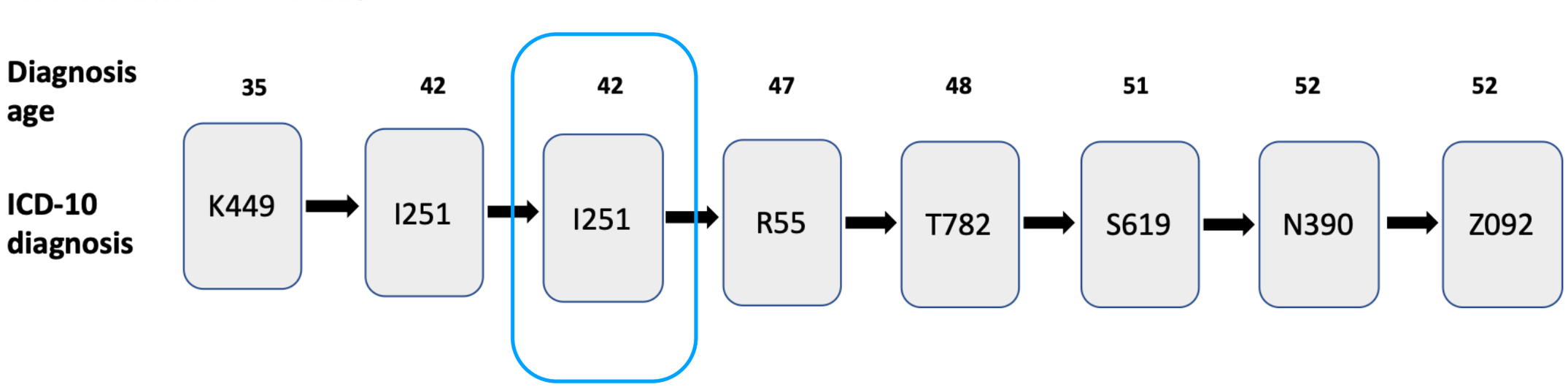
Longitudinal diagnosis history data are available in large data sets

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



Longitudinal diagnosis history data are available in large data sets

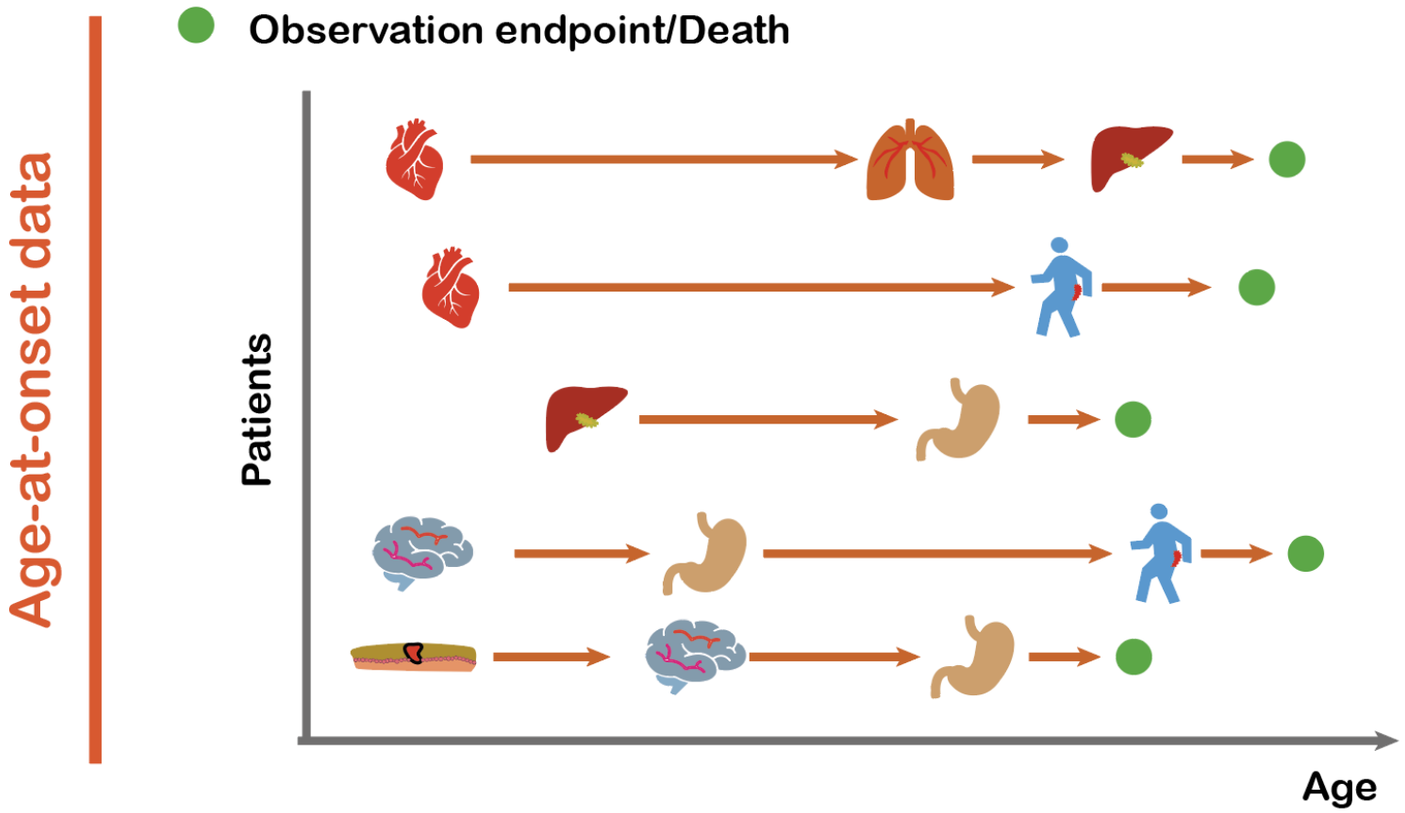
EHR data provides individual-level age-at-onset information



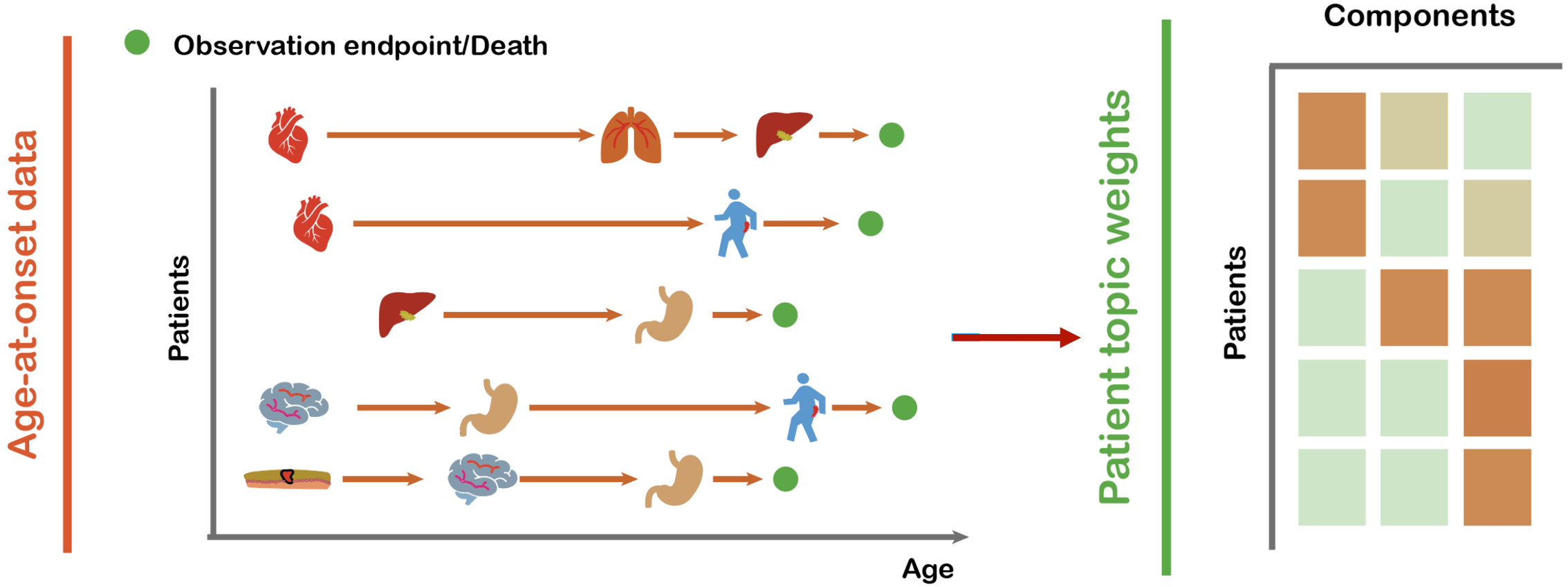
Age-at-onset analysis

- 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

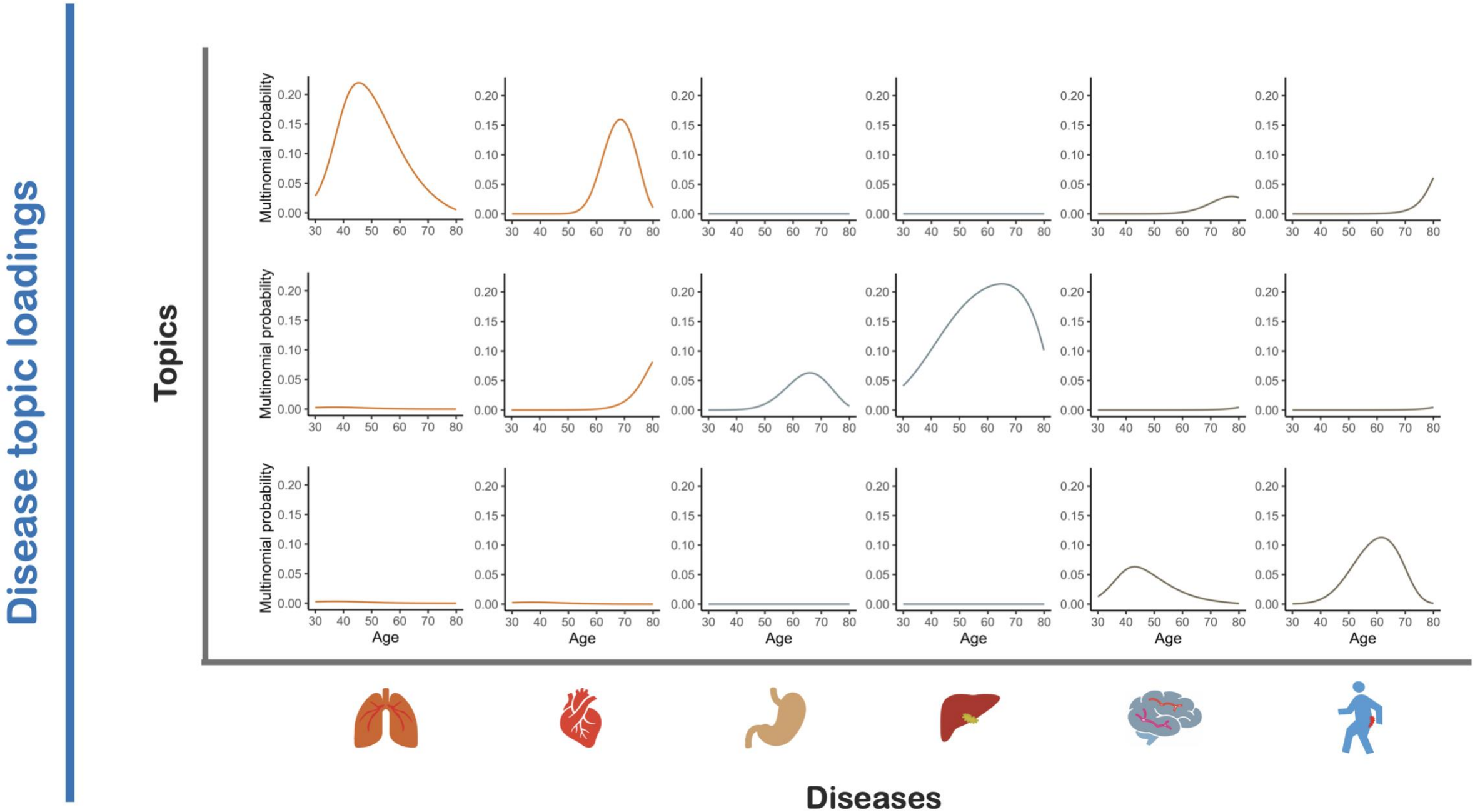
Infer latent comorbidities to represent patient level diagnosis history



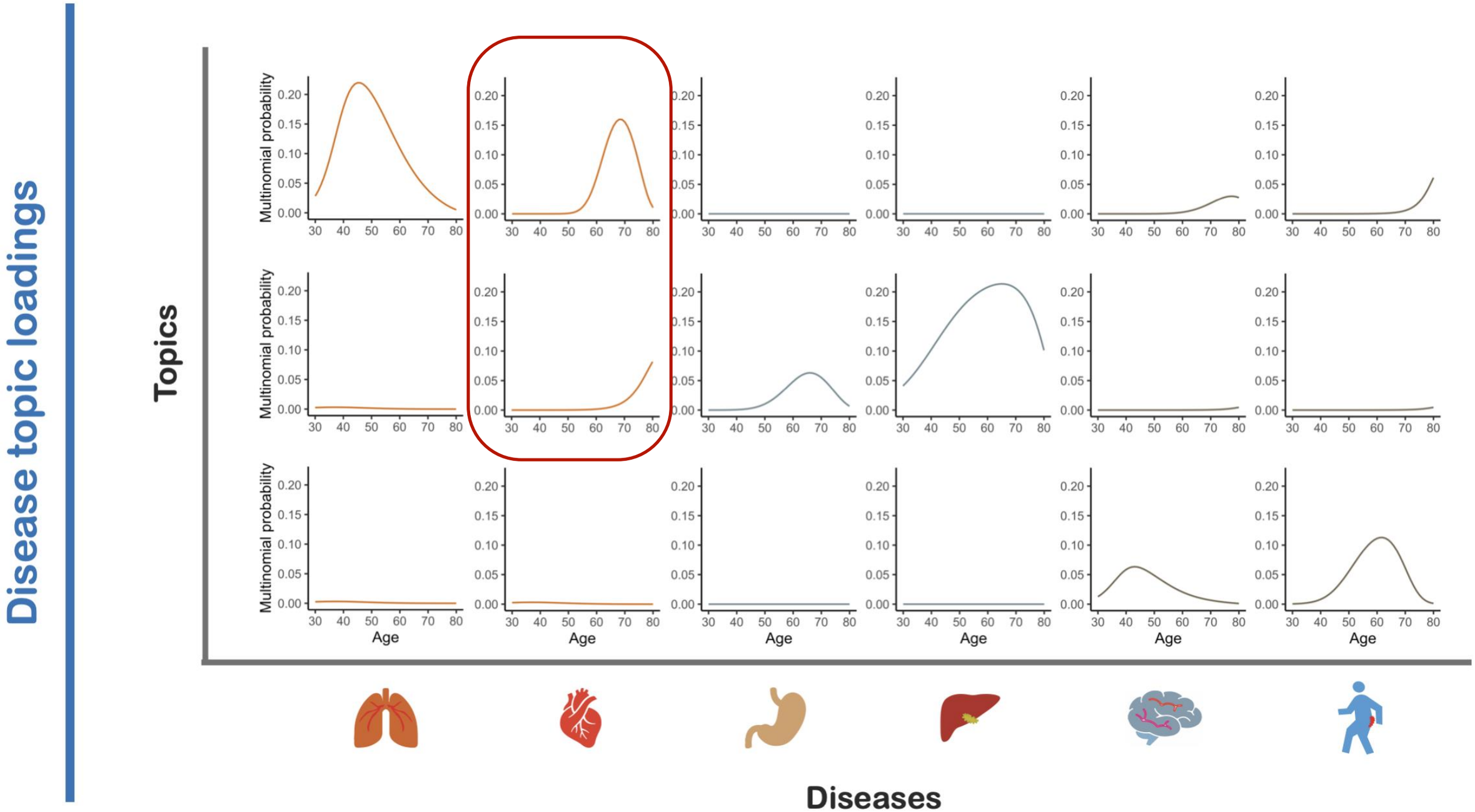
Infer latent comorbidities to represent patient level diagnosis history



Infer latent comorbidities to represent patient level diagnosis history

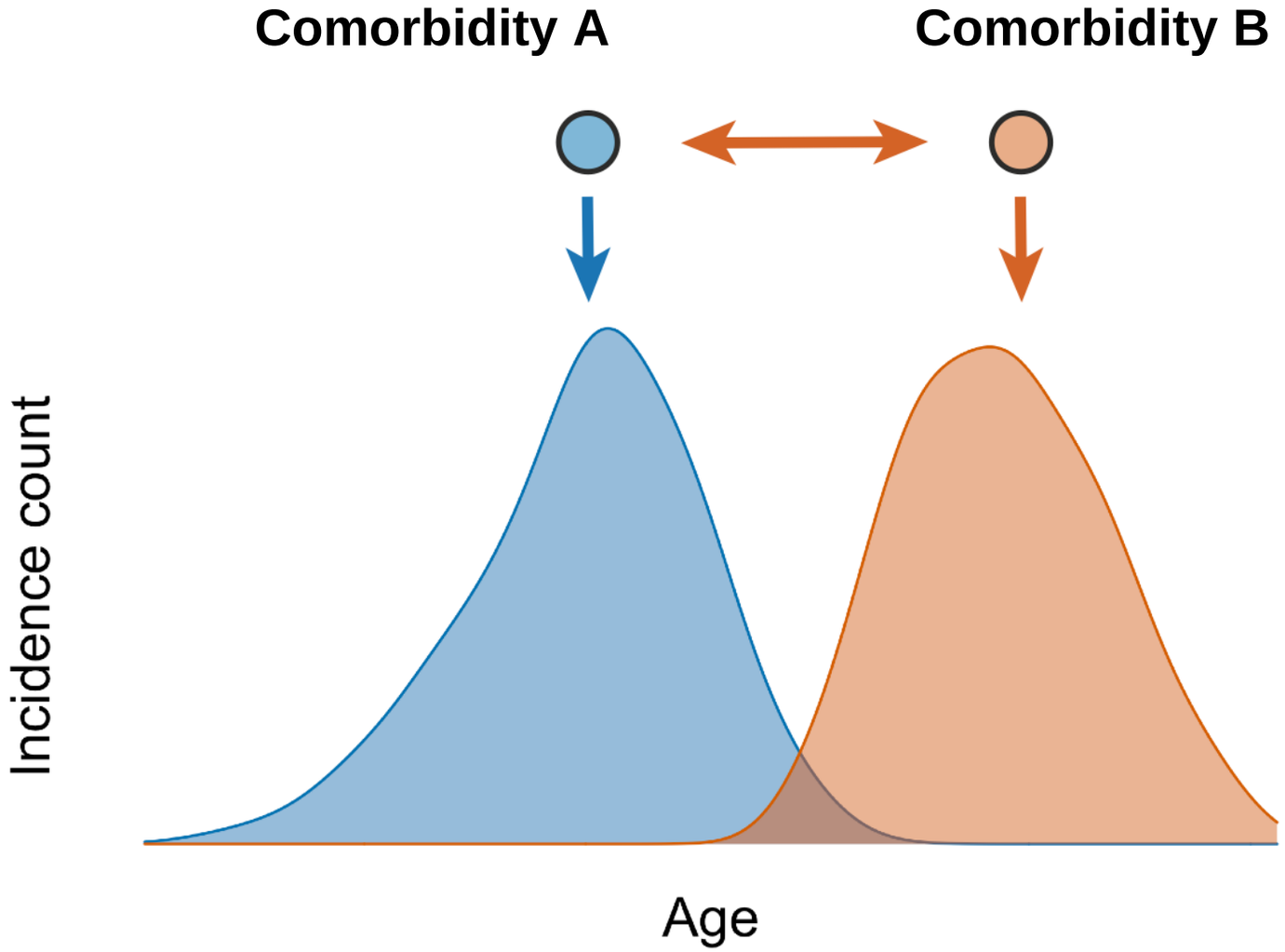


Infer latent comorbidities to represent patient level diagnosis history

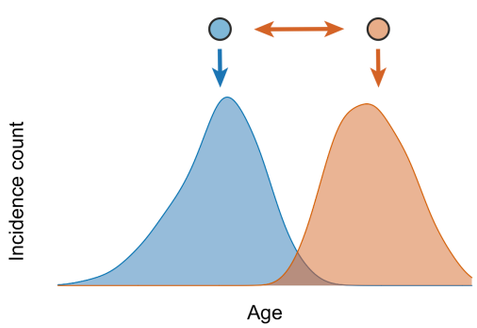


- Background: EHR system collects longitudinal patient medical history
- Methods
  - Age-dependent topic modelling of disease records infer latent comorbidities that represent the full patient diagnosis history
- Results
  - **Identify distinct comorbidity profiles that are associated with different 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

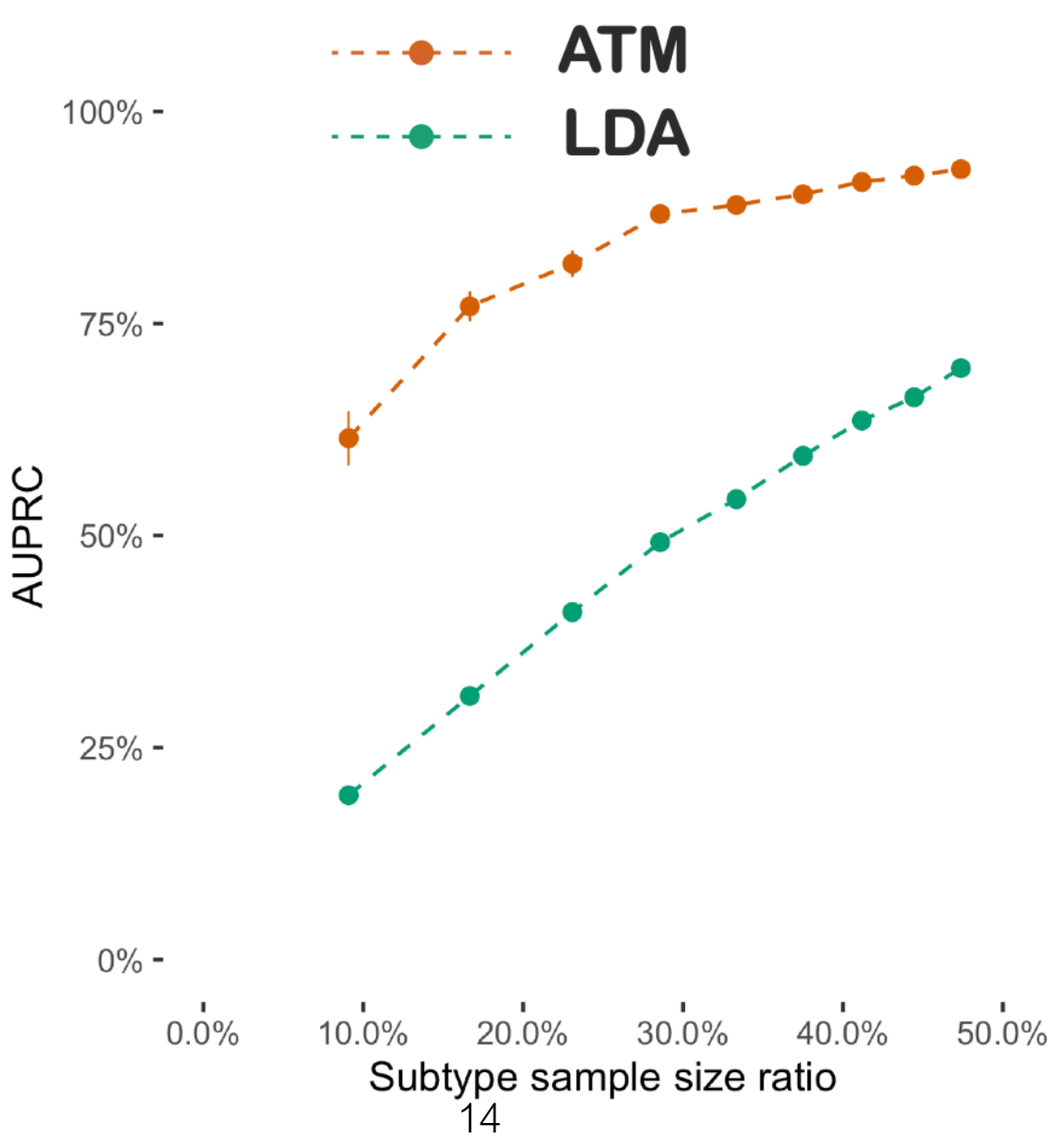
Simulation analysis shows the model could capture age dependent comorbidity



Simulation analysis shows the model could capture age dependent comorbidity



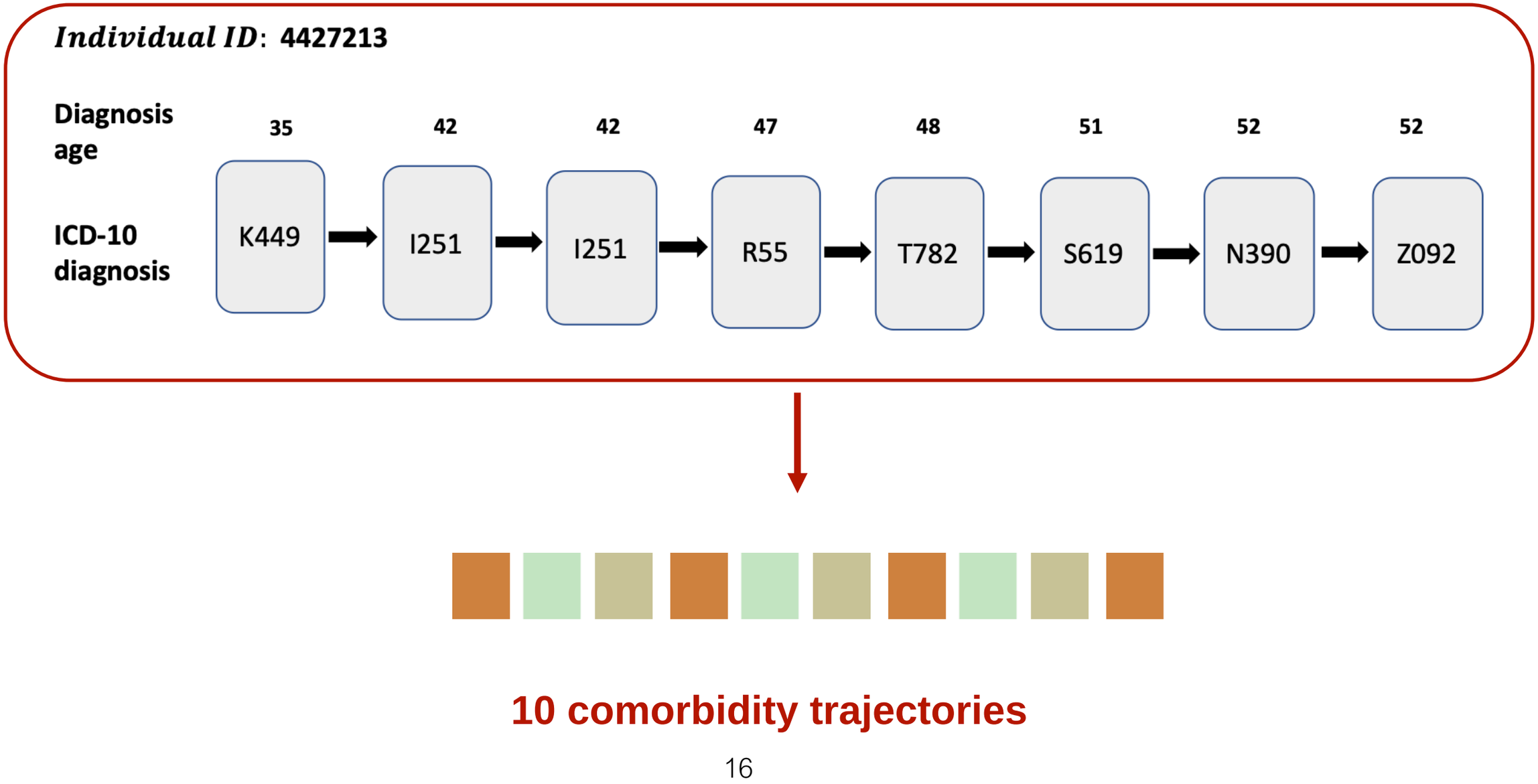
Age-dependent comorbidity modelling (red) versus non-age modelling (green)



- 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

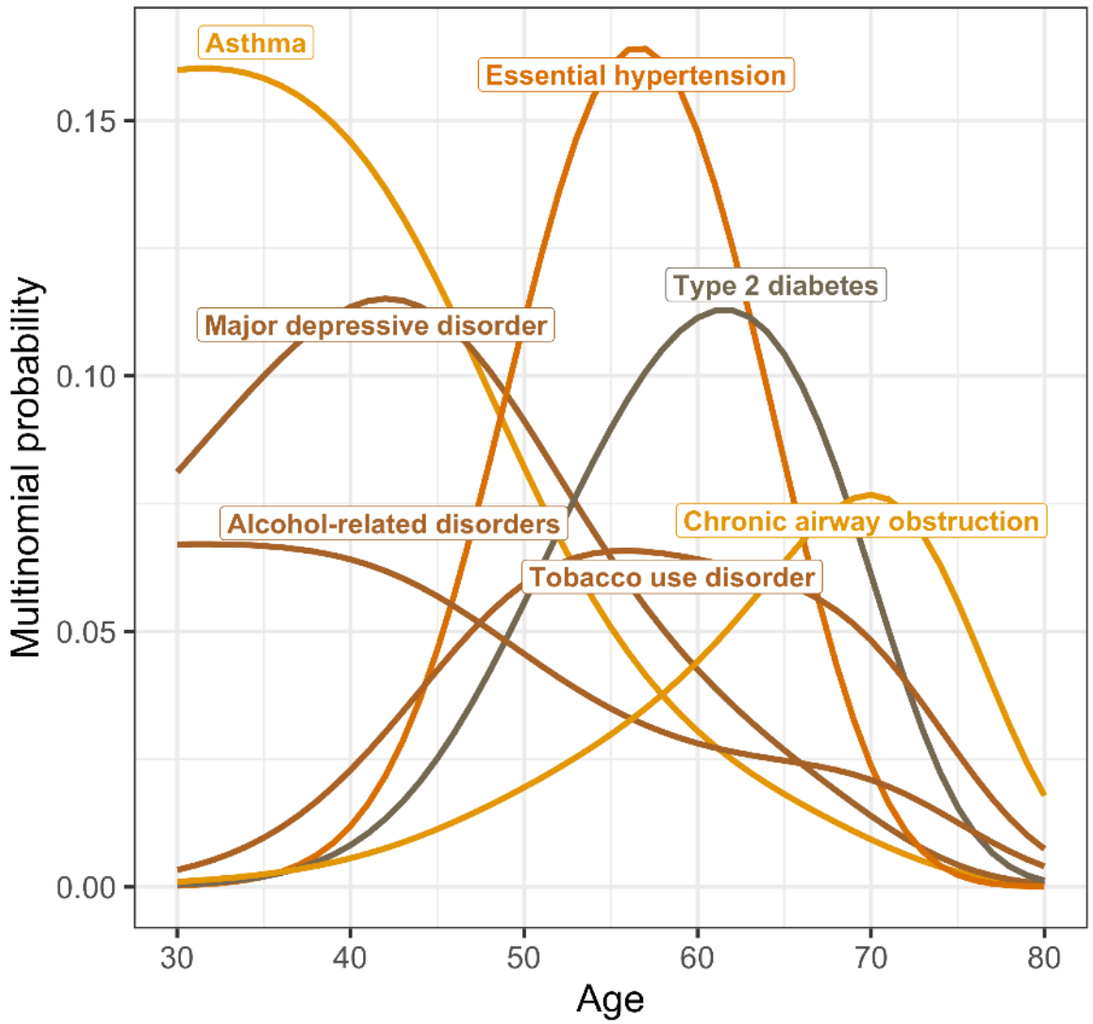
1.7 million records across 348 diseases in the UK Biobank



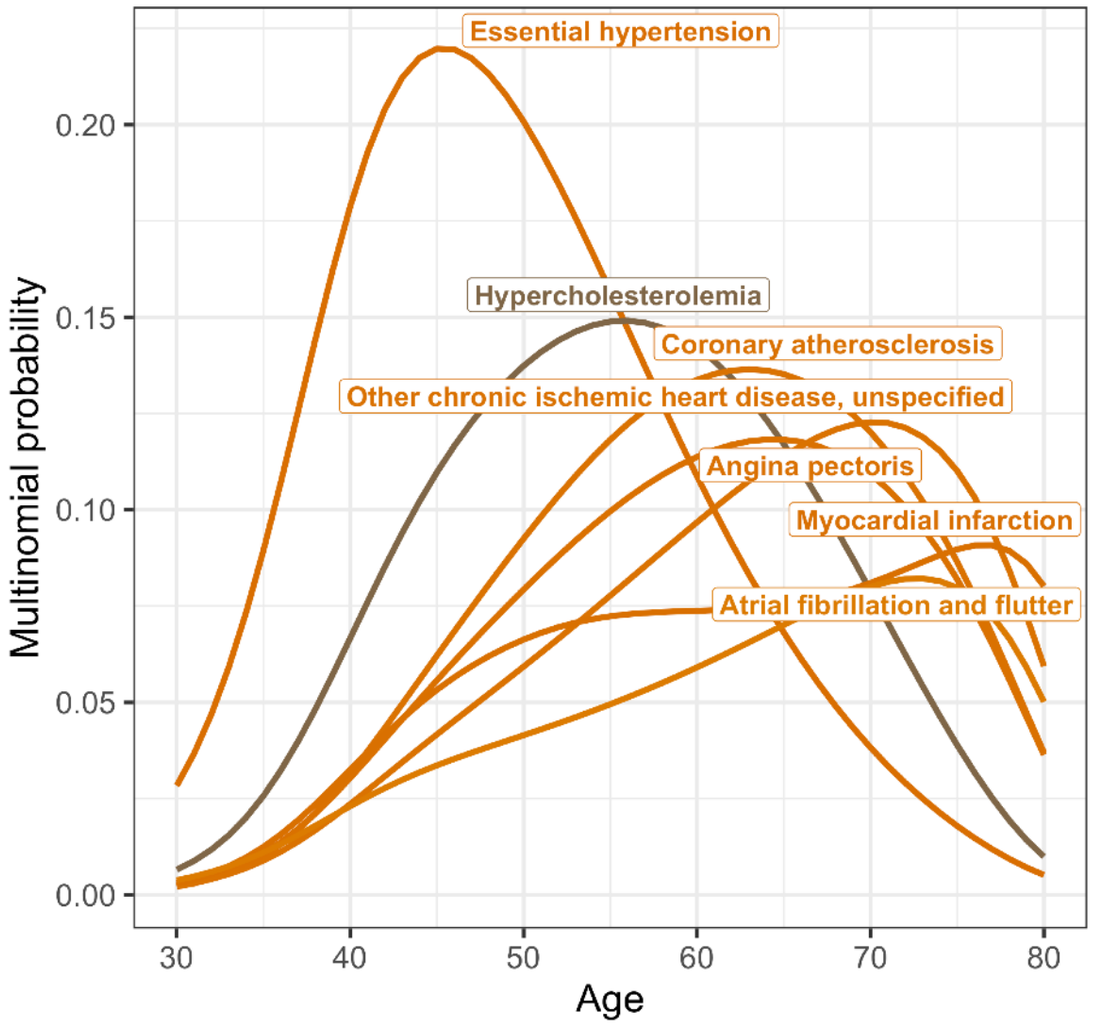
10 comorbidity trajectories

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

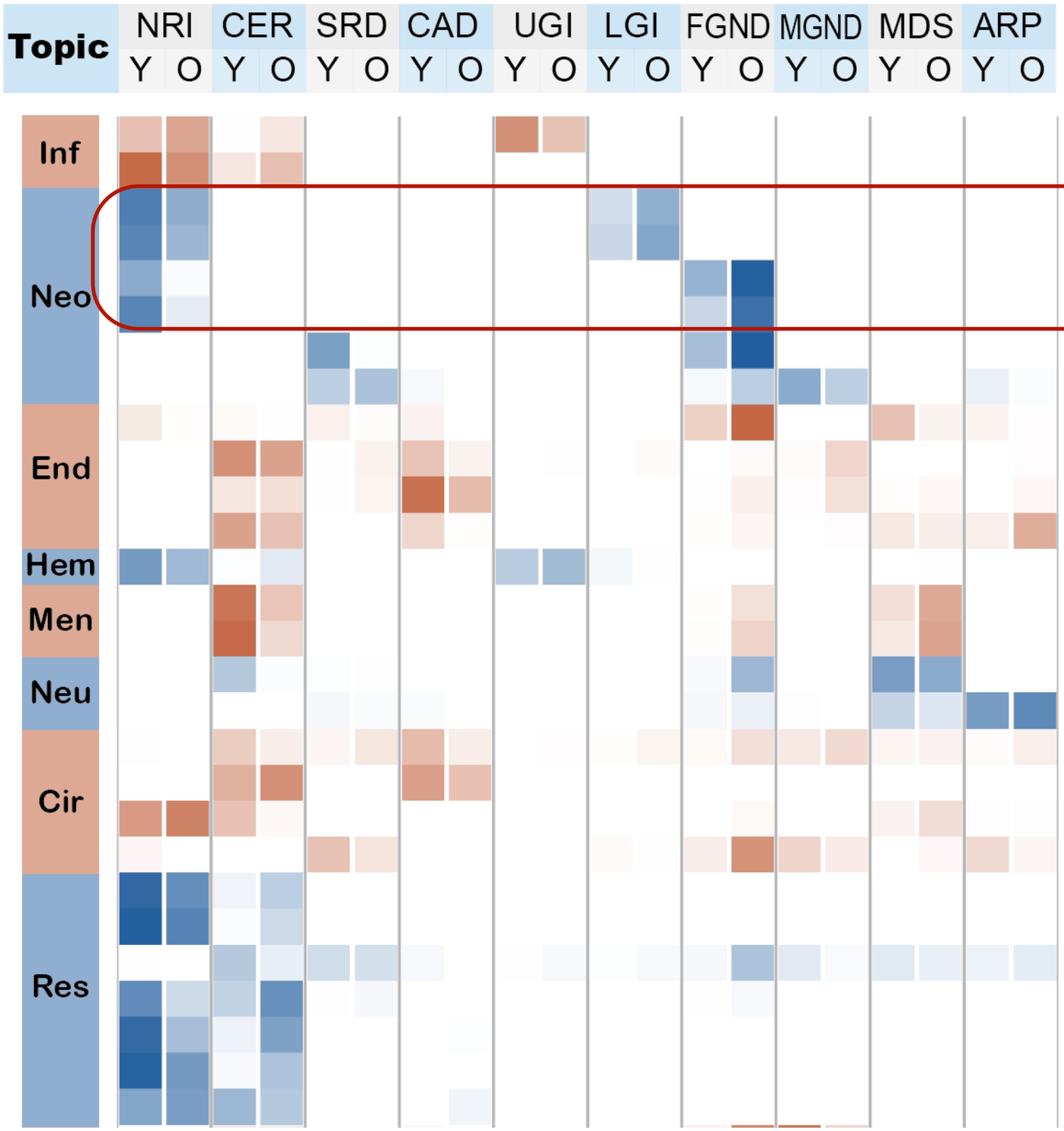
Y **Young group: < 60**

O **Old group: ≥ 60**

Posterior topic weights

0 — 1

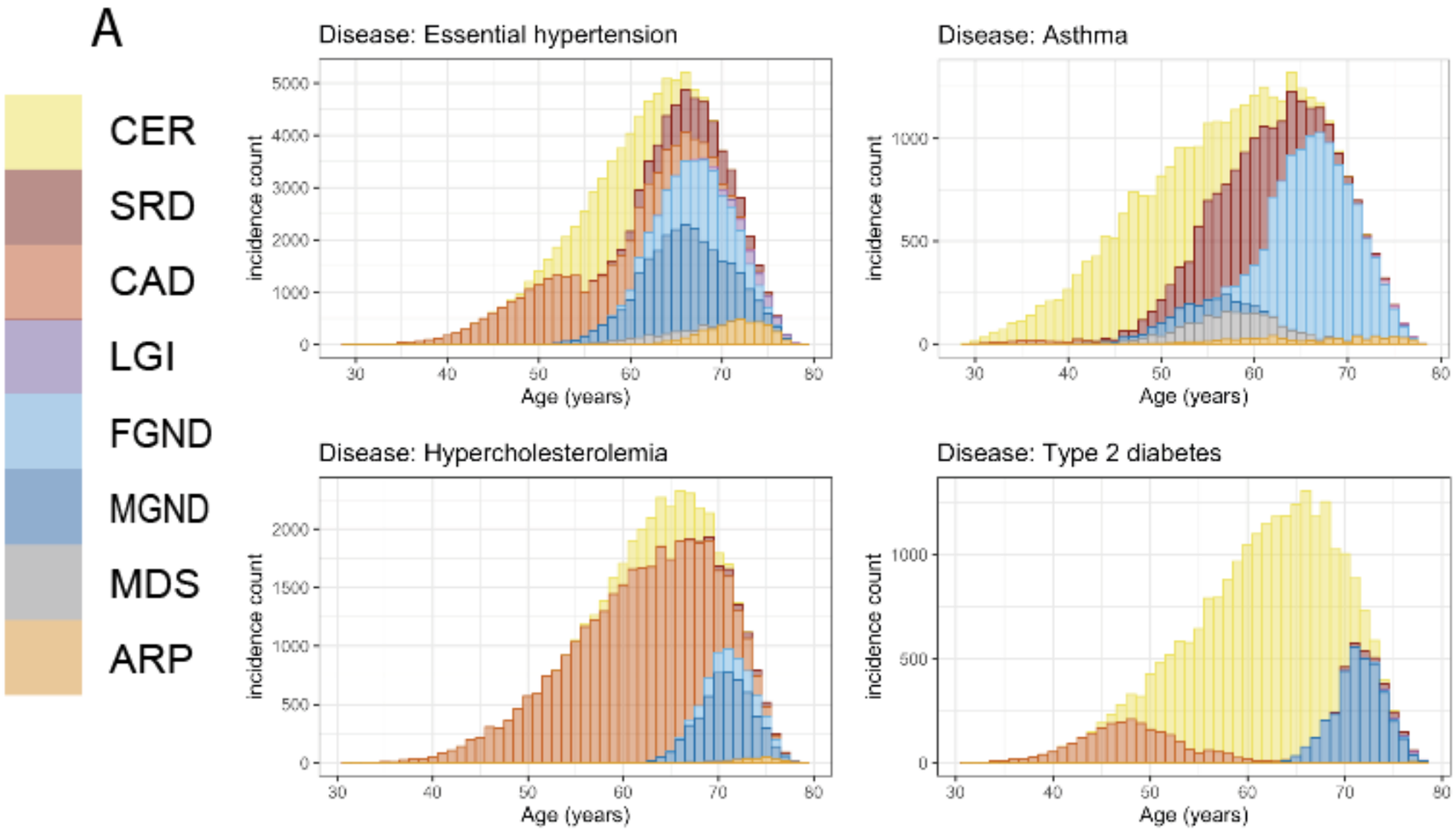
NRI: neoplasm, respiratory, infection
LGI: lower gastrointestinal
FGND: Female genitourinary, neoplasm, digestive



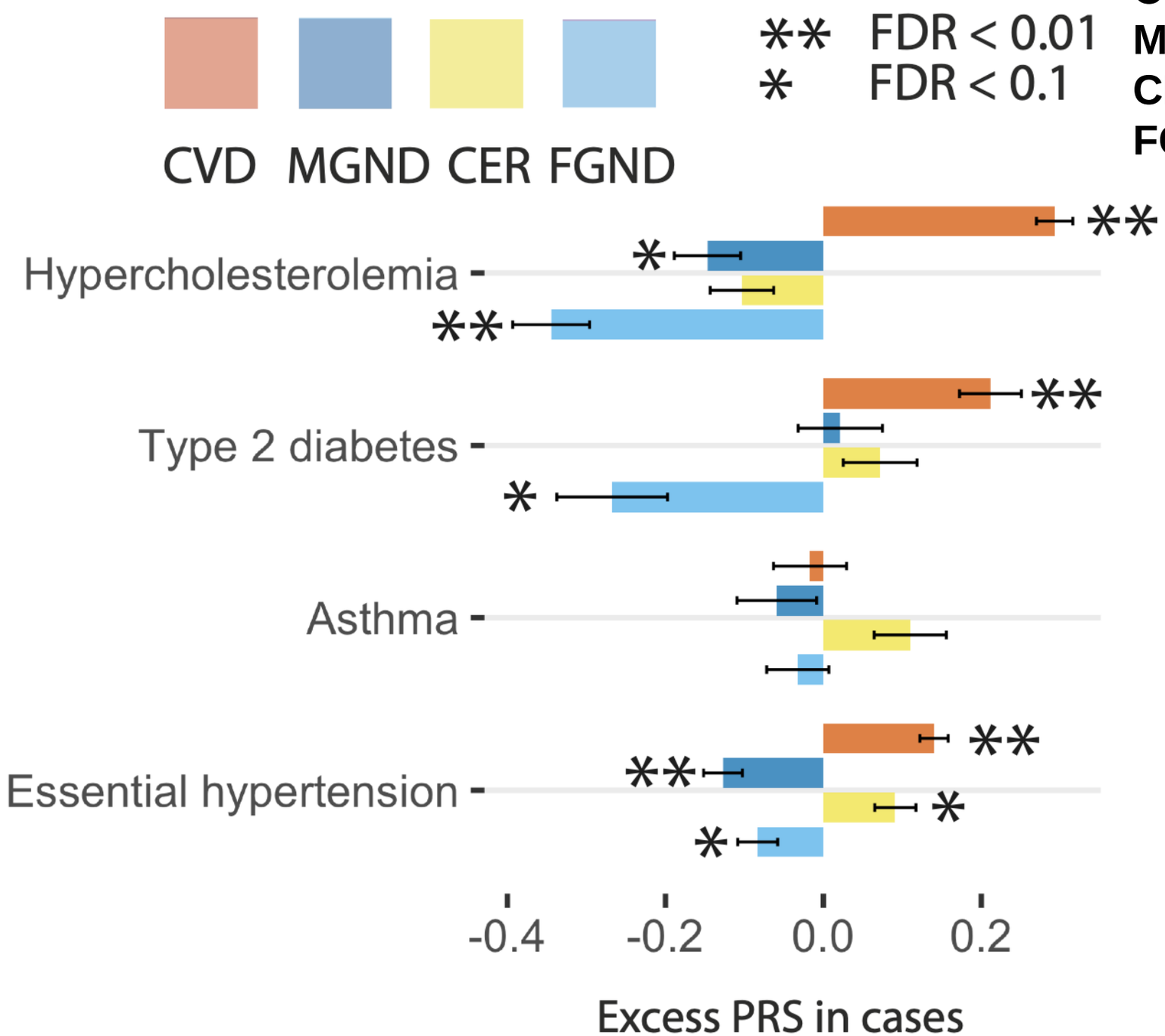
- Background: EHR system collects longitudinal patient medical history
- Methods
  - Age-dependent topic modelling of disease records infer latent comorbidities that represent the full patient diagnosis history
- Results
  - Identify distinct comorbidity profiles that are associated with different age-at-onsets.
  - **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



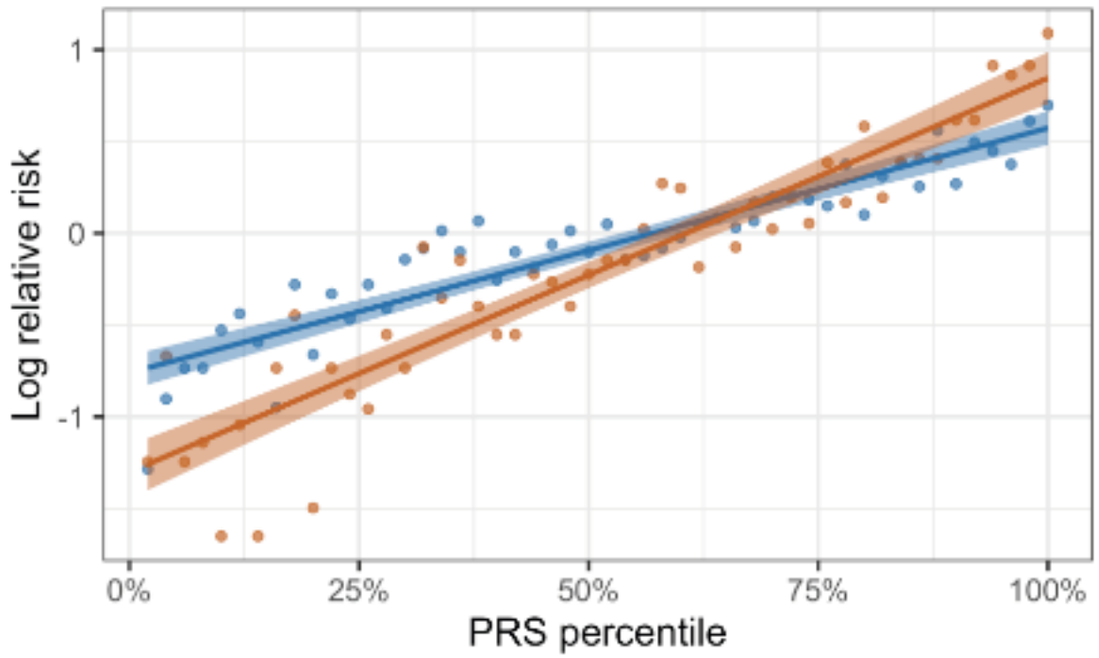
CVD: Cardiovascular disease
MGND: Male genitourinary
CER: Type 2 diabete topic (endocrine)
FGND: Female genitourinary

Disease PRS stratification power varies across disease subtypes

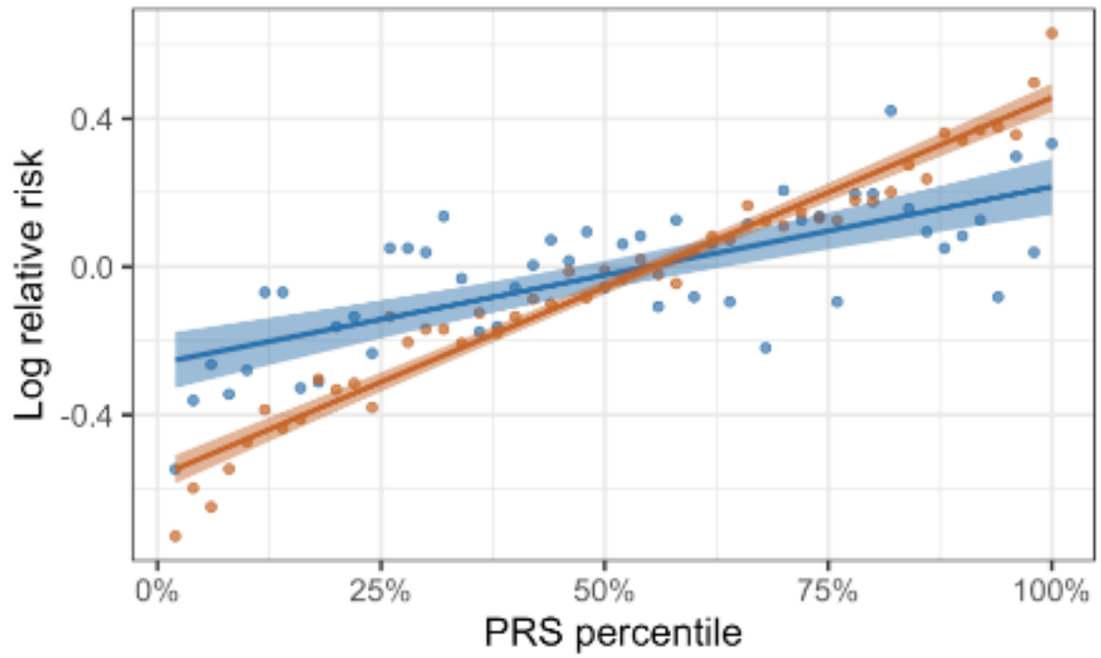
CVD: Cardiovascular disease
MGND: Male genitourinary

CVD MGND CER FGND

Disease: Type 2 diabetes

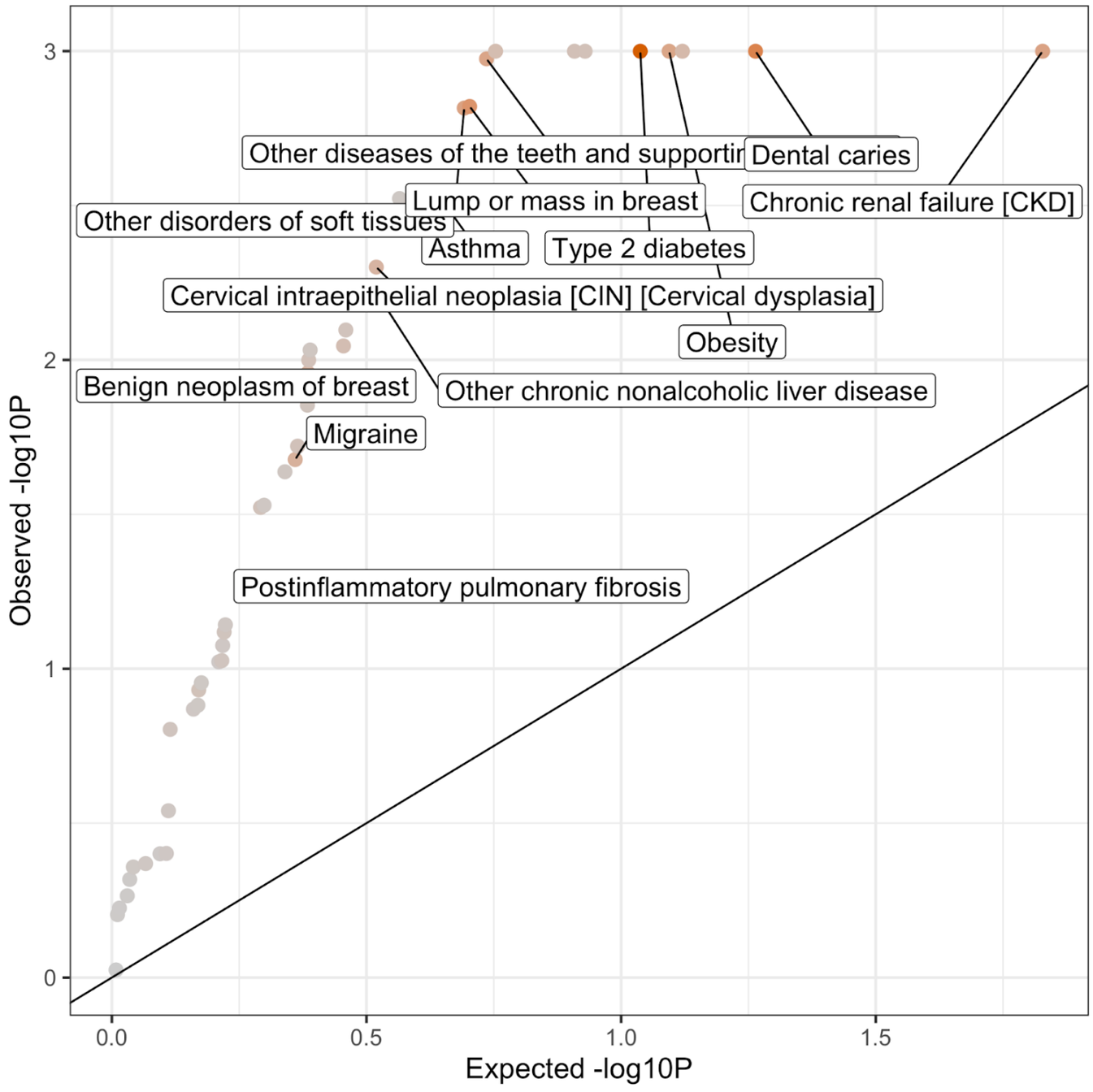


Disease: Hypercholesterolemia



- 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

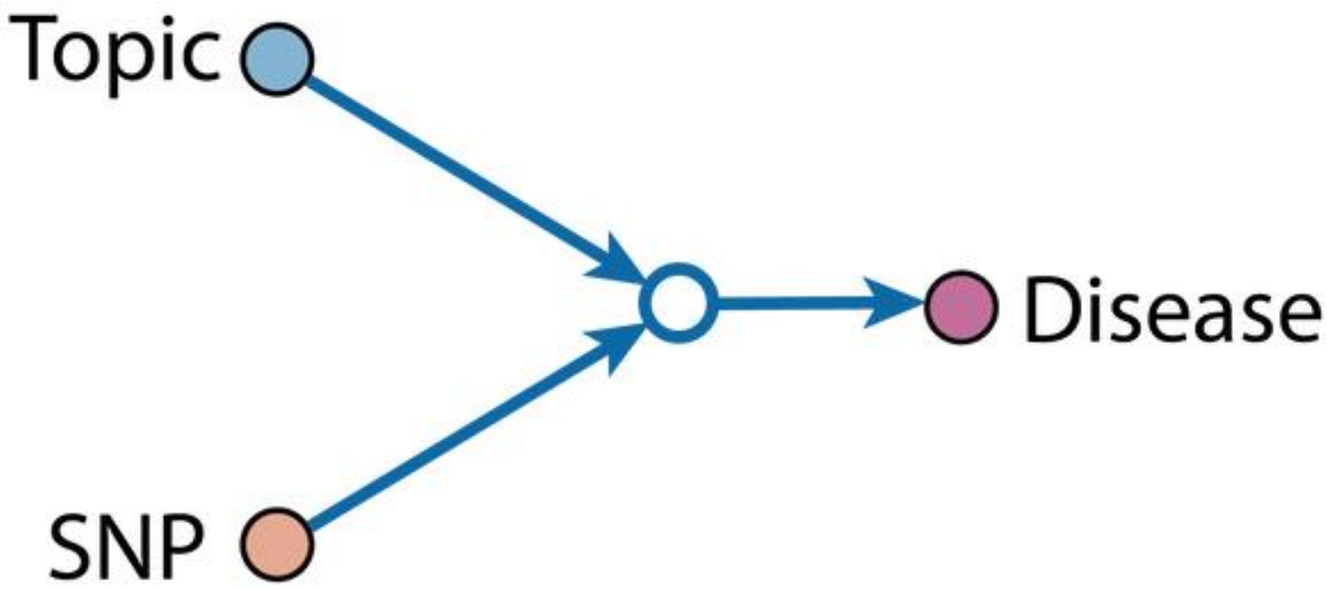
Genetic heterogeneity across subtypes using *Fst* and Genetic correlations



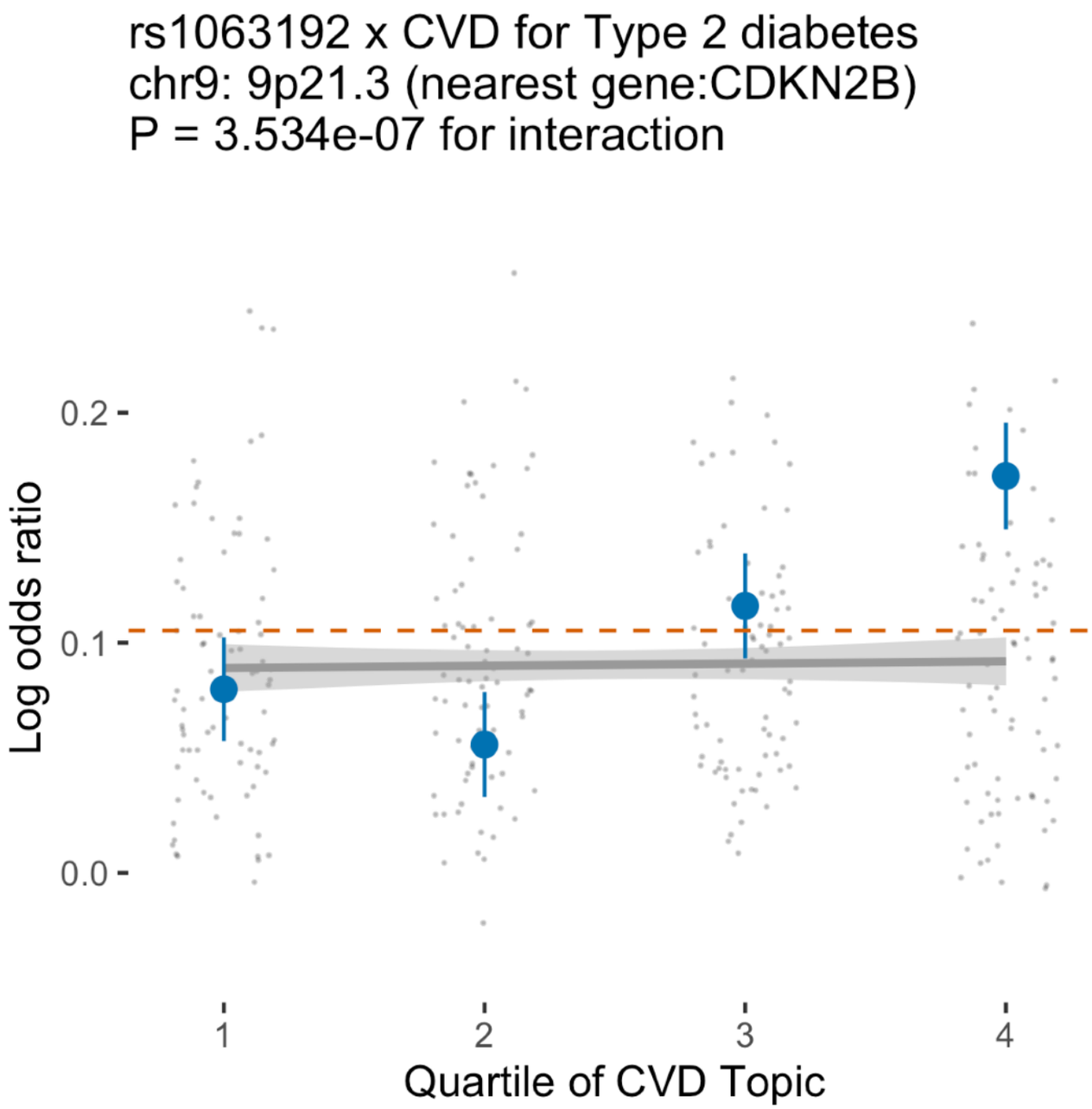
P-values for *Fst* analysis of diseases stratified by topic liabilities. 29 diseases have excessive *Fst* across subtypes identified by Comorbidity.

- 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



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.

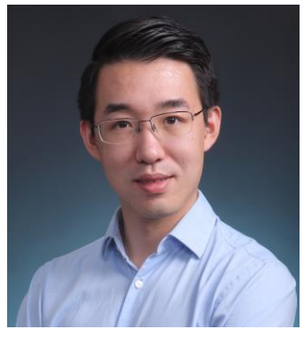
MedRxiv: Age-dependent topic modelling of comorbidities in UK Biobank identifies disease subtypes with differential genetic risk

Acknowledgement:

Gil McVean

Alkes Price

Chris Holmes



Martin Zhang

Yidong Zhang

& Members of Alkes group

Funding:

Wellcome Trust studentship

Rhodes Scholarship

BIG DATA INSTITUTE

UNIVERSITY OF OXFORD

British Heart Foundation

Thank you!

xilinjiang@hsph.harvard.edu