Step 2: Obtain Phenotype Topic Prior

PheCode - ICD Code Mapping

PheCode	Disease	ICD Code
250	Diabetes mellitus	250, ...
495	Asthma	493, ...
495.1	Chronic obstructive asthma	493.XX,...
...	...	...

2-Decimal PheCodes

PheCode Freq.

$\pi \times \alpha$

θ

T, δ

$\bar{z}$ Average topic proportion

$w, h_0(\cdot)$

Parameters for the Survival Model

Survival Information

✗ >20 Days

✓ 5 Days

Predicted Survival Curve

Latent Topic for EHR token

z

EHR TOKEN

x

ϕ

Patient Phenotype Topic Mixture

Step 1: Data PreProcess

Raw EHR Data

Multiple Types of EHR Data

- **ICD Codes**
- **Prescriptions**
- **Lab Tests**
- **Doctor Notes**

Step 3: Model Inference

"Words" distribution in each data type

New Raw EHR Data

Step 4: Prediction

Patient Phenotype Topic Mixture For new patient