# BEHRT - Transformer for Electronic Health Records

Nova Search Reading Group

by: Simão Gonçalves

# Context

Electronic Health Records

# Context

Electronic Health Records

Data:

1. Diagnostics
2. Procedures
3. Lab results
4. Medication
5. Medical notes

# Context - Research Tasks in EHRs

**General tasks:**

1. Medical Concept Extraction
2. Patient Trajectory Modeling
3. Disease Inference

**Tasks in Clinical Notes:**

1. Single concept extraction (e.g. tagging)
2. Temporal Event Extraction
3. Relation Extraction
4. Abbreviation Expansion

# Context - Research Tasks in EHRs

**General tasks:**

1. Medical Concept Extraction
2. Patient Trajectory Modeling
3. Disease Inference

# Challenges

1. Complex and nonlinear interactions among past, present and future concepts.

# Challenges

1. Complex and nonlinear interactions among past, present and future concepts.
2. Long-term dependencies among concepts

# Challenges

1. Complex and nonlinear interactions among past, present and future concepts.
2. Long-term dependencies among concepts
3. Difficulties of representing multiple heterogeneous concepts of variable sizes to the model

# Challenges

1. Complex and nonlinear interactions among past, present and future concepts.
2. Long-term dependencies among concepts
3. Difficulties of representing multiple heterogeneous concepts of variable sizes to the model
4. Irregular intervals between consecutive visits.

# BEHRT - Introduction

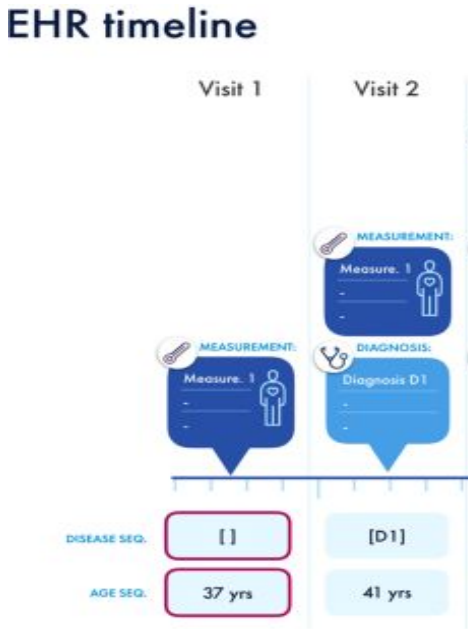
Let each diagnose be a word, each visit a sentence, and a patient’s entire medical history a document.

# BEHRT - Data Prep

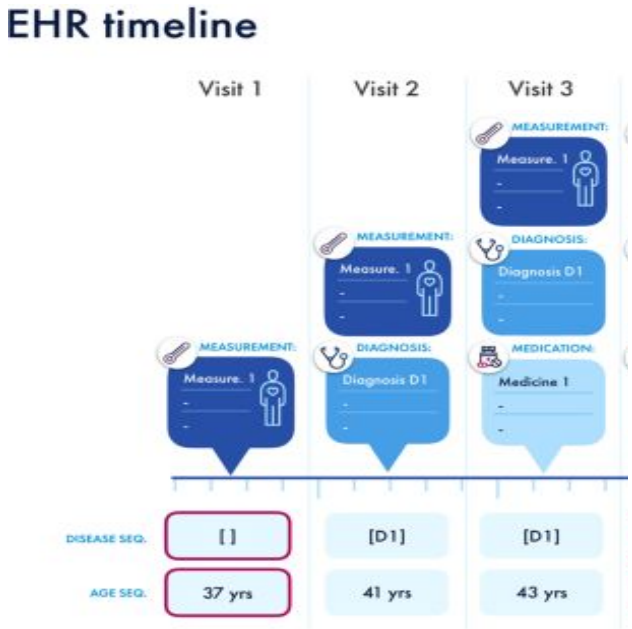
Data Source: Clinical Practice Research Datalink (CPRD)

# BEHRT - Data Prep

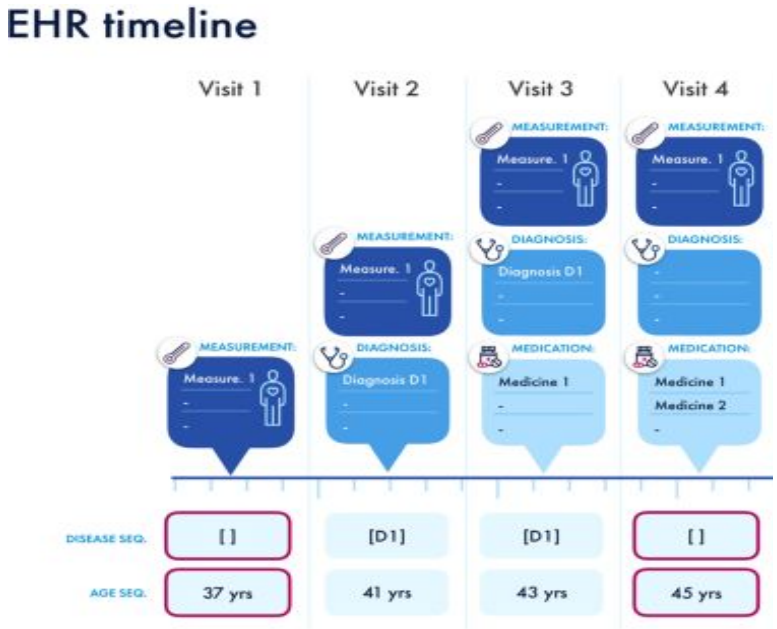
# BEHRT - Data Prep



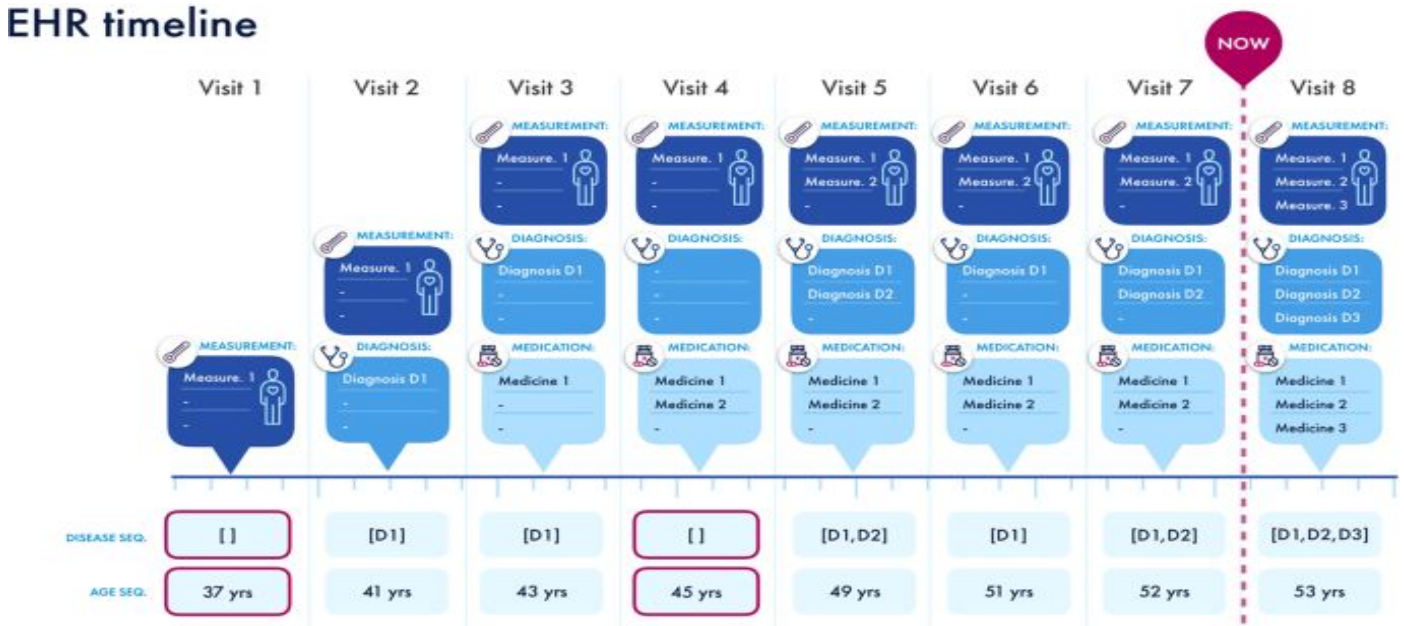
# BEHRT - Data Prep



# BEHRT - Data Prep



# BEHRT - Data Prep

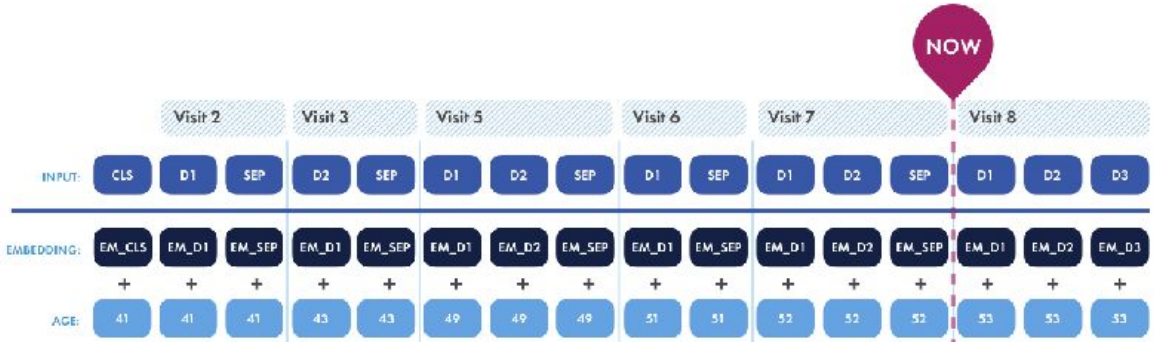


# BEHRT - Modeling

1. Encode diagnosis along the hospital visits

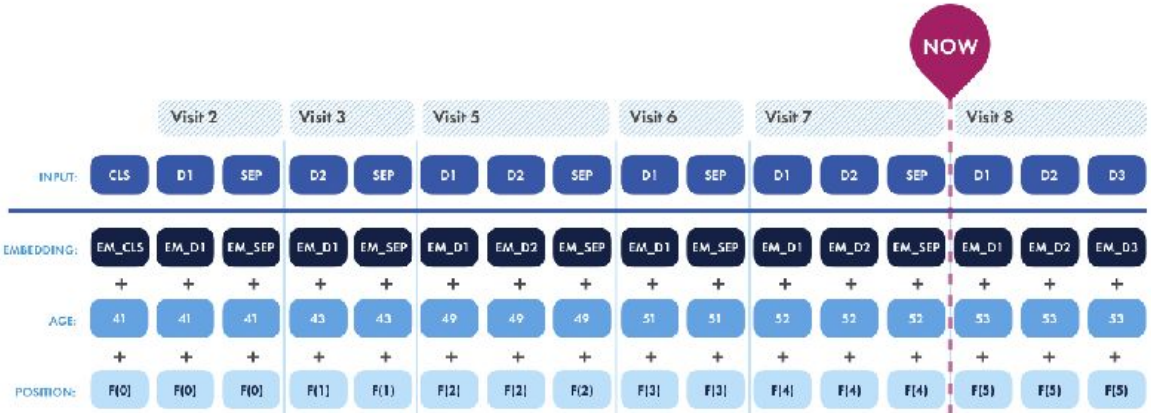
# BEHRT - Modeling

1. Encode diagnosis along the hospital visits
2. Add patient age: encodes time + relation between diseases and age



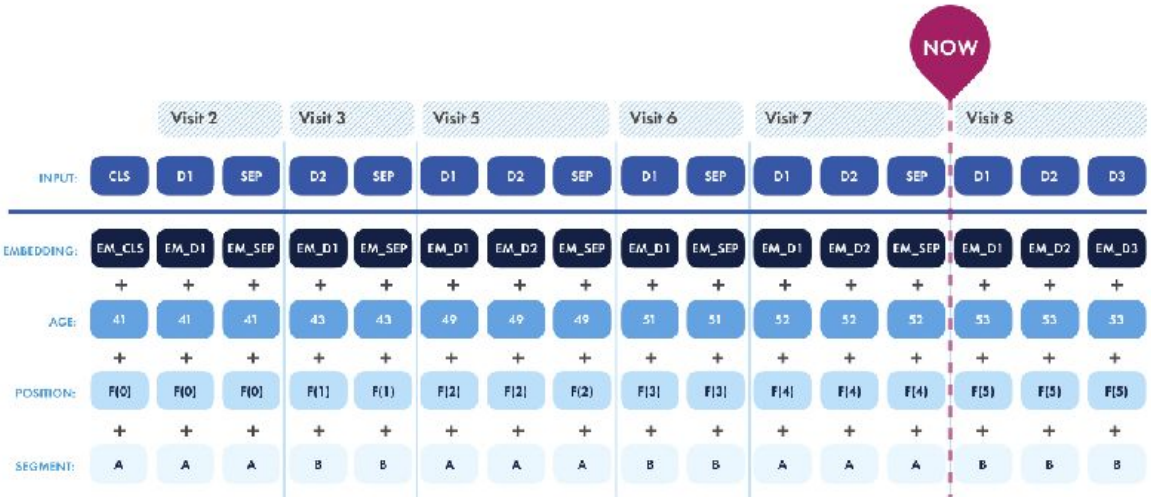
# BEHRT - Modeling

1. Encode diagnosis along the hospital visits
2. Add patient age: encodes time + relation between diseases and age
3. Add Position: catch positional interactions among diseases



# BEHRT - Modeling

1. Encode diagnosis along the hospital visits
2. Add patient age: encodes time + relation between diseases and age
3. Add Position: catch positional interactions among diseases
4. Add Segment: give extra information about separation between visits



# BEHRT - Modeling

1. Encode diagnosis along the hospital visits
2. Add patient age: encodes time + relation between diseases and age
3. Add Position: catch positional interactions among diseases
4. Add Segment: give extra information about separation between visits
5. Final Embedding

# BEHRT - Why use BERT architecture

**Advantage of Bidirectional self awareness**:

1. Order which diseases appear can be reversed
2. Time interval between two diagnoses can vary

# BEHRT - Why use BERT architecture

**Advantage of Bidirectional self awareness**:

1. Order which diseases appear can be reversed
2. Time interval between two diagnoses can vary

These are caused by:

A. Variabilities in individuals' health

B. Practice of care

C. Random events

# BEHRT - Why use BERT architecture

**Advantage of Bidirectional self awareness**:

1. Order which diseases appear can be reversed
2. Time interval between two diagnoses can vary

These are caused by:

A. Variabilities in individuals' health

B. Practice of care

C. Random events

Also, **BERT architecture is suited for the challenges we mentioned earlier**

# BEHRT - Why use BERT architecture

**Advantage of Bidirectional self awareness**:

1. Order which diseases appear can be reversed
2. Time interval between two diagnoses can vary

These are caused by:

A. Variabilities in individuals' health

B. Practice of care

C. Random events

Also, **BERT architecture is suited for the challenges we mentioned earlier**:

1. Complex and nonlinear interactions in time.

# BEHRT - Why use BERT architecture

**Advantage of Bidirectional self awareness**:

1. Order which diseases appear can be reversed
2. Time interval between two diagnoses can vary

These are caused by:

A. Variabilities in individuals' health

B. Practice of care

C. Random events

Also, **BERT architecture is suited for the challenges we mentioned earlier**:

1. Complex and nonlinear interactions in time.
2. Long-term dependencies among concepts

# BEHRT - Why use BERT architecture

**Advantage of Bidirectional self awareness**:

1. Order which diseases appear can be reversed
2. Time interval between two diagnoses can vary

These are caused by:

A. Variabilities in individuals' health

B. Practice of care

C. Random events

Also, **BERT architecture is suited for the challenges we mentioned earlier**:

1. Complex and nonlinear interactions in time.
2. Long-term dependencies among concepts
3. Difficulties of representing multiple heterogeneous concepts of variable sizes

# BEHRT - Why use BERT architecture

**Advantage of Bidirectional self awareness**:

1. Order which diseases appear can be reversed
2. Time interval between two diagnoses can vary

These are caused by:

A. Variabilities in individuals' health

B. Practice of care

C. Random events

Also, **BERT architecture is suited for the challenges we mentioned earlier**:

1. Complex and nonlinear interactions in time.
2. Long-term dependencies among concepts
3. Difficulties of representing multiple heterogeneous concepts of variable sizes
4. Irregular intervals between consecutive visits.

# BEHRT - Training

Standard BERT training:

1. Mask 12% of words
2. Replace 1.5% of words with noise
3. Perform Masked LM and update parameters

# BEHRT - Learning

3 tasks, predict:

1. Next visit's diseases
2. Diseases in next 6 months
3. Diseases in next 12 months

# BEHRT - Learning

3 tasks, predict:

1. Next visit's diseases
2. Diseases in next 6 months
3. Diseases in next 12 months

Data selection:

1. train test split (80-20) at patient level
2. Task 2 and 3 must have 6/12 months of visits after the 4th visit.

# BEHRT - Learning

3 tasks, predict:

1. Next visit's diseases (699k patients)
2. Diseases in next 6 months (391k patients)
3. Diseases in next 12 months (342 patients)

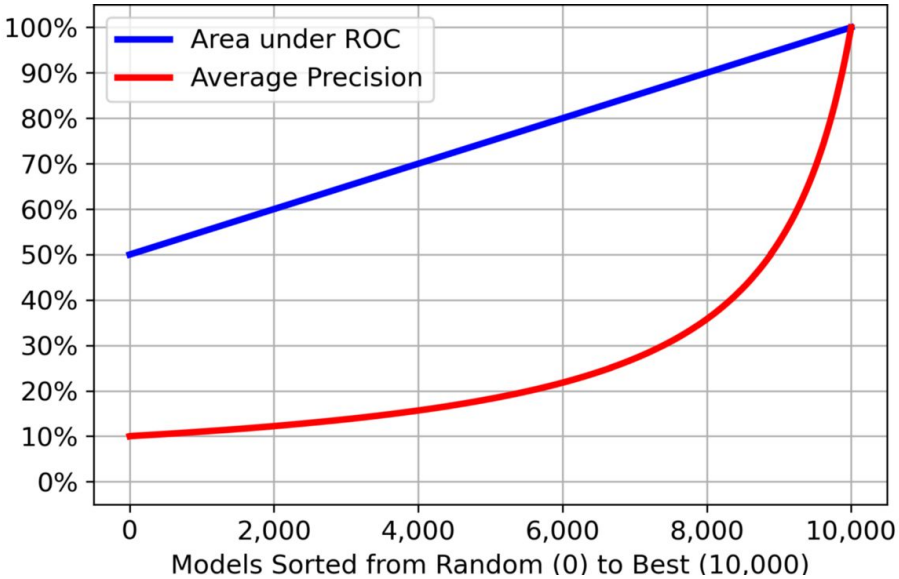
Data selection:

1. train test split (80-20) at patient level
2. Task 2 and 3 must have visits 6/12 months after the 4th visit.

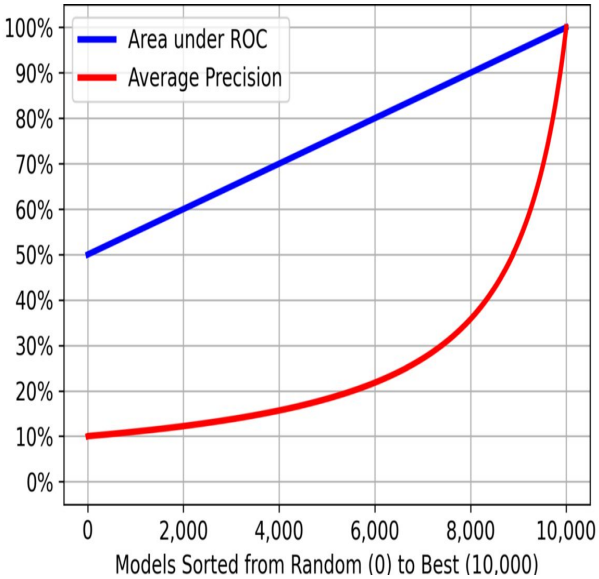
# BEHRT - Results

| Model Name | Next Visit (APS\|AUROC) | Next 6 M (APS\|AUROC) | Next 12 M (APS\|AUROC) |
|---|---|---|---|
| **BEHRT** | **0.462\|0.954** | **0.525\|0.958** | **0.506\|0.955** |
| Deepr | 0.360\|0.942 | 0.393\|0.943 | 0.393\|0.943 |
| RETAIN | 0.382\|0.921 | 0.417\|0.927 | 0.413\|0.928 |

# BEHRT - Results (Average Precision vs AUC-ROC)

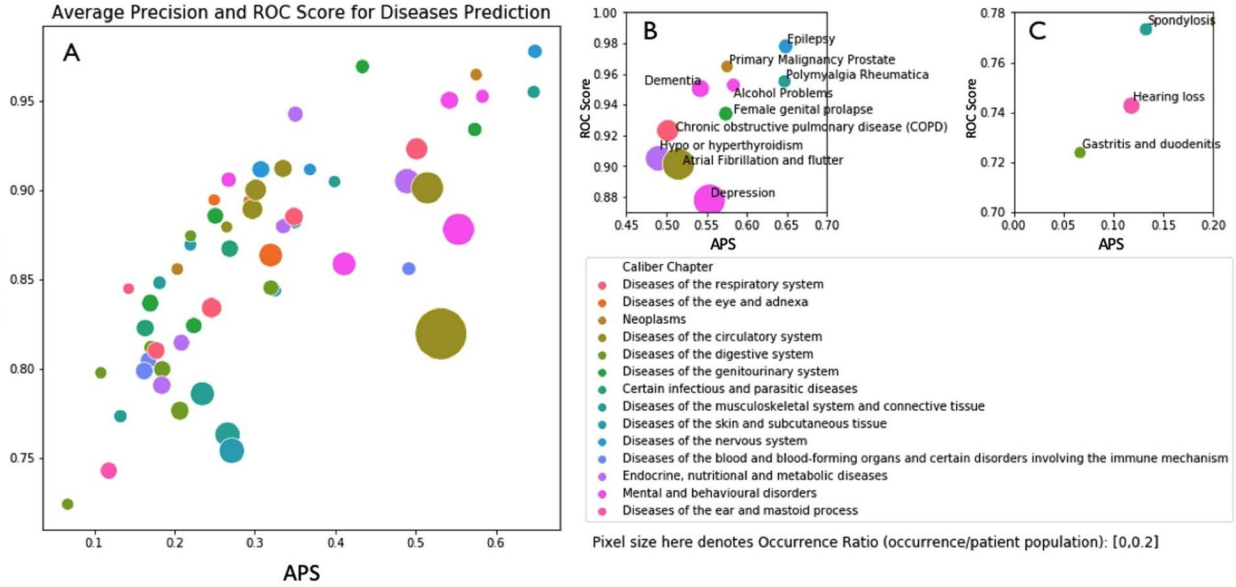


# BEHRT - Results (Average Precision vs AUC-ROC)



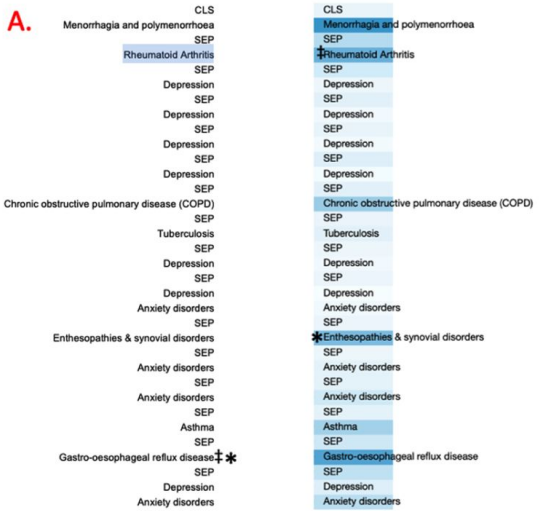

| Model Name | Next Visit (APS\|AUROC) | Next 6 M (APS\|AUROC) | Next 12 M (APS\|AUROC) |
|---|---|---|---|
| **BEHRT** | **0.462\|0.954** | **0.525\|0.958** | **0.506\|0.955** |
| Deepr | 0.360\|0.942 | 0.393\|0.943 | 0.393\|0.943 |
| RETAIN | 0.382\|0.921 | 0.417\|0.927 | 0.413\|0.928 |

# BEHRT - Interpretability




Pixel size here denotes Occurrence Ratio (occurrence/patient population): [0,0.2]

# BEHRT - Interpretability

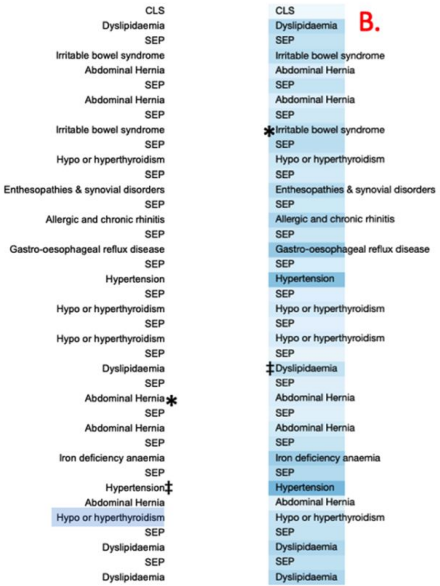



In A, BEHRT shows strong relationships between:

1. Rheumatoid Arthritis and Menorrhagia and polymenorrhoea
2. Rheumatoid Arthritis and Enthesopathies & synovial Disorders
3. Gastro-oesophageal reflux disease and Rheumatoid Arthritis (not shown‡)
4. Gastro-oesophageal reflux disease and Enthesopathies & synovial Disorders (not shown*)

In B, BEHRT shows strong relationships between:

1. Hypo or Hyperthyroidism and Hypertension
2. Hypo or Hyperthyroidism and Gastro-oesophageal reflux disease
3. Hypertension and Dyslipidaemia (not shown‡)
4. Abdominal Hernia and Irritable Bowel Syndrome (not shown*)



# BEHRT - Interpretability (gender)

**Table S7.** Predictions of Gender-Specific Diseases (Test Predictions from Next 6 Months)

| Disease Name | Disease Gender | Male Predictions | Female Predictions |
|---|---|---|---|
| Hyperplasia of Prostate | M | 384 | 0 |
| Hydrocoele (incl infected) | M | 36 | 0 |
| Male Infertility | M | 1 | 24 |
| Primary Malignancy Prostate | M | 557 | 0 |
| Erectile Dysfunction | M | 425 | 1 |
| Menorrhagia and Polymenorrhoea | F | 0 | 697 |
| Endometriosis | F | 0 | 47 |
| Female Genital Prolapse | F | 0 | 865 |
| Female Infertility | F | 2 | 36 |
| Benign Neoplasm of Ovary | F | 0 | 69 |
| Postmenopausal Bleeding | F | 0 | 140 |
| Primary Malignancy Breast | F | 0 | 11 |
| Primary Malignancy Ovarian | F | 1 | 193 |

Male (M) and Female (F) Predictions imply label predictions >0.5

# BEHRT - Interpretability

**Population Statistics**

Table S1: Statistics of Cohorts Selected Prediction Tasks

| Characteristics | | Next Visit | Next 6 M | Next 12M |
|---|---|---|---|---|
| Gender | Male | 41.80% | 42.30% | 41.70% |
| | Female | 58.20% | 57.70% | 58.30% |
| Ethnicity | White | 46.40% | 48.30% | 47.40% |
| | Unknown | 43.80% | 44.00% | 44.50% |
| | Indian | 0.40% | 0.50% | 0.50% |
| | Other | 0.30% | 0.30% | 0.30% |
| | Pakistani | 0.20% | 0.30% | 0.20% |
| | Black Carib | 0.20% | 0.30% | 0.20% |
| | Other Asian | 0.10% | 0.10% | 0.10% |
| | Black African | 0.10% | 0.10% | 0.10% |
| | Mixed | 0.10% | 0.10% | 0.10% |
| | Bangladeshi | 0.08% | 0.07% | 0.07% |
| | Black Other | 0.07% | 0.06% | 0.06% |
| | Chinese | 0.06% | 0.06% | 0.05% |
| Age Start | 0.25 Quantile | 45 | 46 | 46 |
| | 0.5 Quantile | 58 | 60 | 59 |
| | 0.75 Quantile | 70 | 71 | 70 |
| Age End | 0.25 Quantile | 56 | 58 | 58 |
| | 0.5 Quantile | 70 | 71 | 71 |
| | 0.75 Quantile | 81 | 82 | 82 |
| Unique Codes | 0.25 Quantile | 7 | 8 | 8 |

# Takes and future steps

1. Shows potential to provide value for health
2. Treatments, prescriptions, notes not used here
3. The vectors and attention weights can be used for research
4. These datasets are difficult to access and there aren’t a lot of them