Research Directions using Text Mining on the Stockholm Electronic Patient Record Corpus

Maria Skeppstedt

mariask@dsv.su.se

Health Records in Swedish encompassing 1 Million Patients

Example of a Patient Record

Patient Record nr X:

Anamnesis: The patient says… (free text)

Status: The patient has… (free text)

Assessment: No signs of… (free text)

Gender: Woman
Age: 78
ICD-10: J20.9 (Acute bronchitis)
…

Maria Skeppstedt, DSV

Example of text

Current Status:

Eye movement not assessable, does not open eyes. Movement normal. Babinski's sign missing both sides. Otherwise normal.

Maria Skeppstedt, DSV

Research Directions

1. Diagnose Code Suggestion
2. Triggers for Adverse Events
3. Finding Similar Patient Cases
4. Hypothesis Generation
5. Synthetic Health Records

Maria Skeppstedt, DSV

1) Diagnose Code Suggestion

Example of Code Suggestion

Record nr X

Assessment:

Cough text, text text text text sore throat ....

ICD-10:
J20.9 (Acute bronchitis)

Record nr Y

Assessment:

Text text text, text cough text text, text text ....

ICD-10:
J18.9 (Pneumonia)

Record nr Z

Assessment:

Cough text text, text text sore throat, text text …

ICD-10:
??????

Maria Skeppstedt, DSV

# 2) Triggers for Adverse Events

Example of a Trigger

Trigger 18, T18: Abrupt Cessation of Medication

...so she was switched to Soma. She started to have some confusion and pains <T18>so the Soma was discontinued</T18> and her cognition cleared....
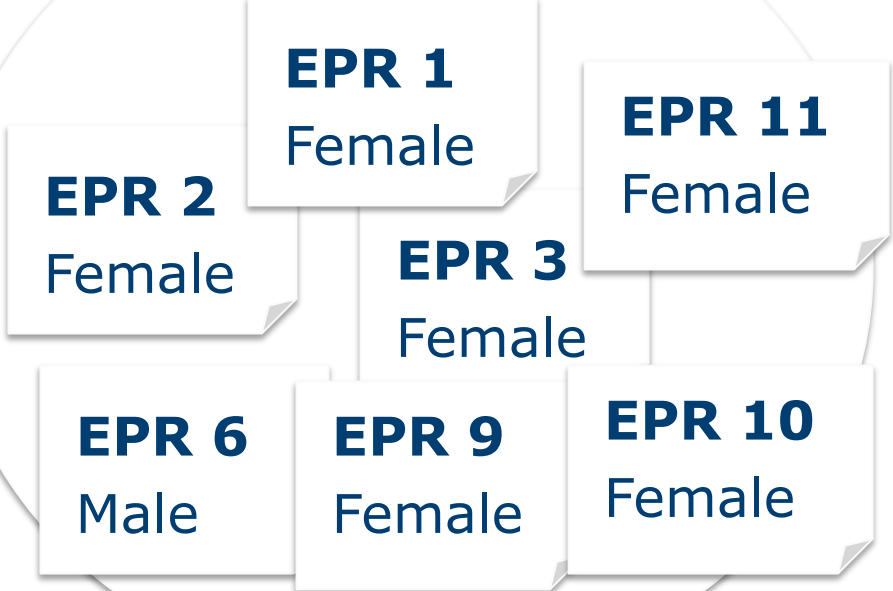
Maria Skeppstedt, DSV

3) Finding Similar Patient Cases

Finding similar patient cases

EPR X

Disease a

Disease a

Disease a

Maria Skeppstedt, DSV
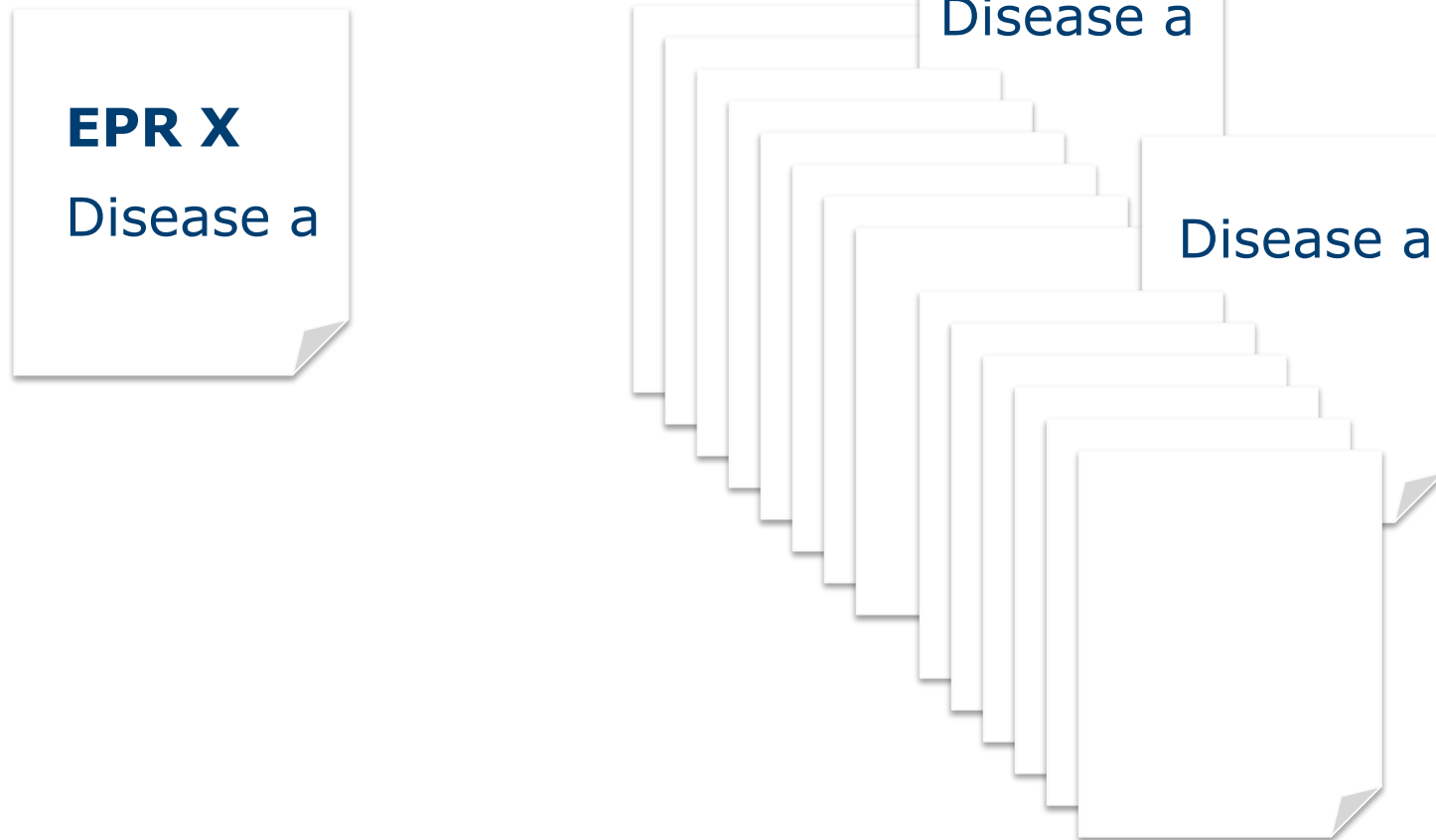
4) Hypothesis Generation

Example of Hypothesis Generation

Another cluster

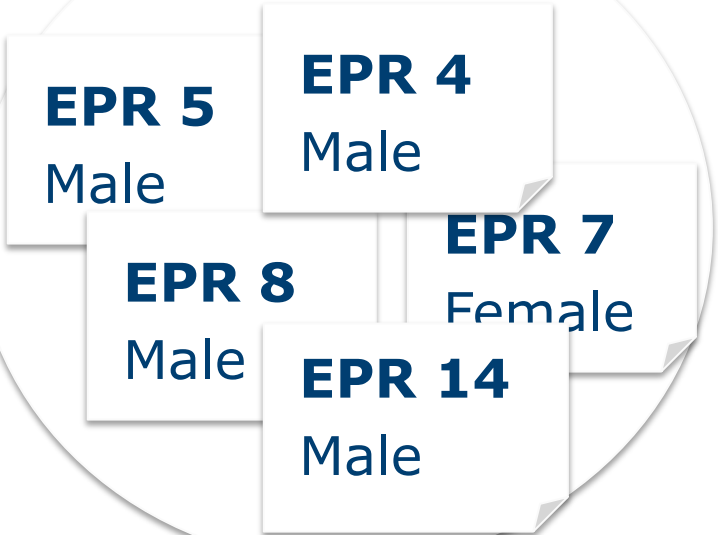Cluster of EPR:s where the patient has *brittleness of the bones*

EPR 13
Male

EPR 1
Male

EPR 15
Male

EPR 1
Female

EPR 2
Female

EPR 11
Female

EPR 3
Female

EPR 6
Male

EPR 9
Female

EPR 10
Female

Another cluster

EPR 5
Male

EPR 4
Male

EPR 8
Male

EPR 7
Female

EPR 14
Male

Hip fracture, cane, …

Maria Skeppstedt, DSV

Stockholm University
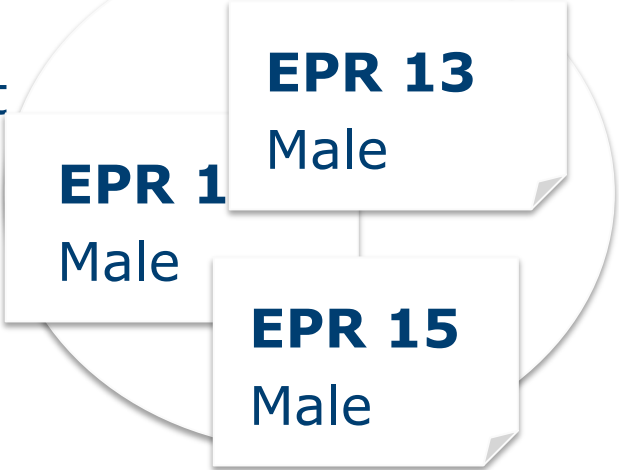
5) Synthetic Health Records

Synthetic Record Plan, Example

No of records: 100

Age: 55-75

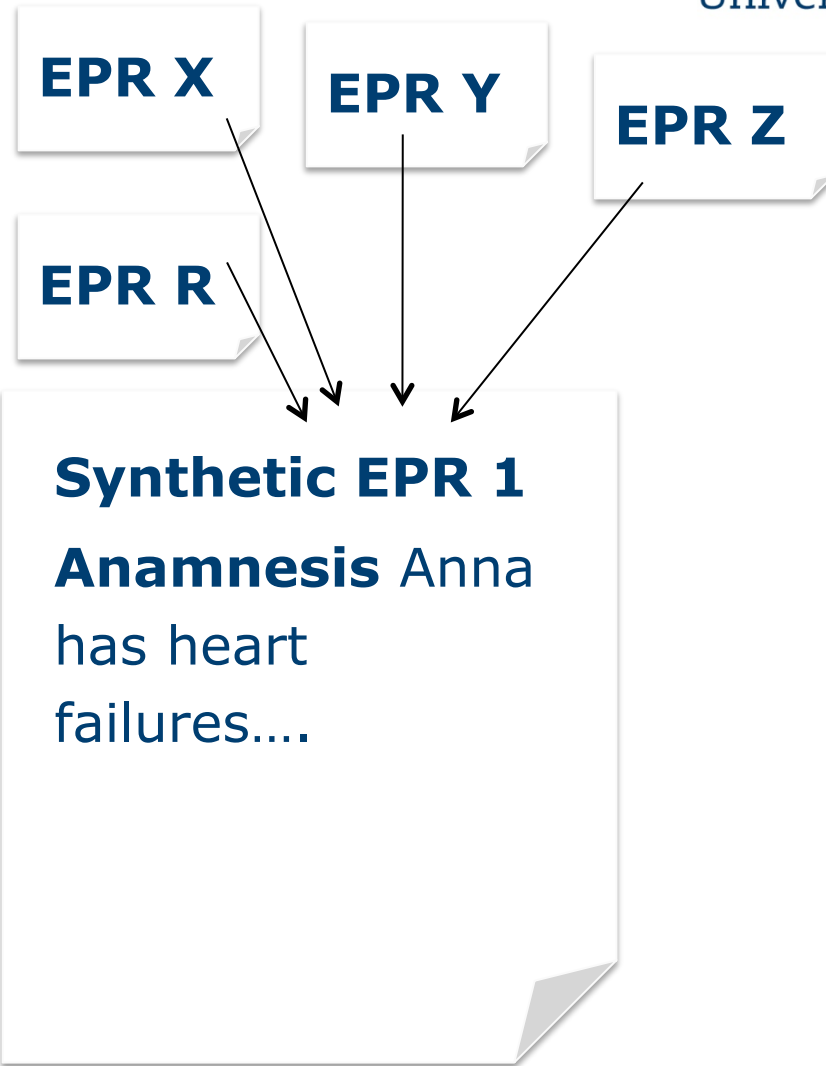**Time period:**

20080101-20080302

ICD-10:

Z853, M059, R104, Z450

Min record size:

100 words

Max record size:

200 words

Maria Skeppstedt, DSV

EPR X

EPR Y

EPR Z

EPR R

Synthetic EPR 1

Anamnesis Anna has heart failures….

Summary

1. Diagnose Code Suggestion
2. Triggers for Adverse Events
3. Finding Similar Patient Cases
4. Hypothesis Generation
5. Synthetic Health Records

Maria Skeppstedt, DSV

Questions?

Maria Skeppstedt

mariask@dsv.su.se