UNIVERSITY OF
CAMBRIDGE

ICUnity: A software tool to harmonise the MIMIC-III and AmsterdamUMCdb databases

Emma Rocheteau[1,2], Jacob Deasy[2], Luca Filipe Roggeveen[3] and Dr Ari Ercole[4]

[1] School of Clinical Medicine, University of Cambridge; [2] Department of Computer Science and Technology, University of Cambridge; [3] Amsterdam University Medical Centre; [4] Division of Anaesthesia, Department of Medicine, University of Cambridge

Problems of working with EHRs

- The quality of raw data is variable, creating significant pre-processing burden
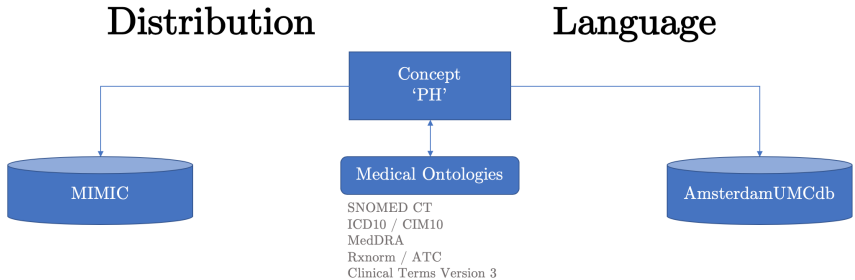
- Lack of standardisation between datasets

Can we (partially) automate data mapping and feature pre-processing between clinical databases?

Specific Challenges
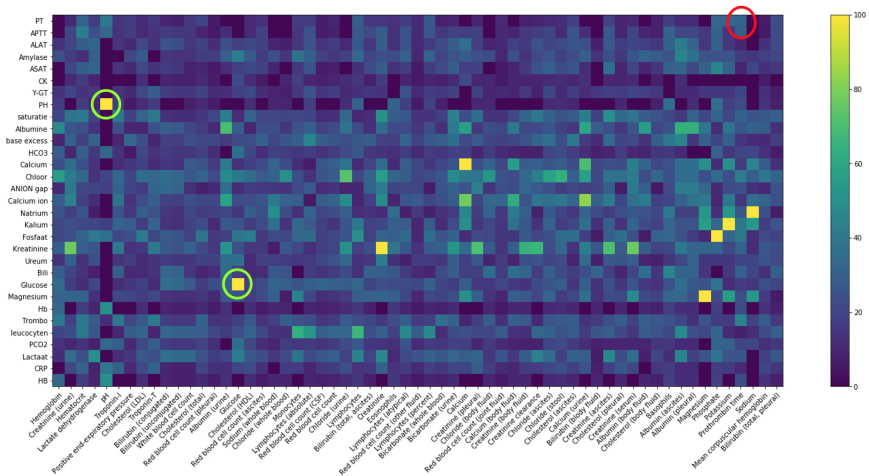
Problems	Examples
Different formats	MIMIC, eICU, AmsterdamUMCdb
Multiple & different label IDs	"HR", "heartrate", "pulse"
Different languages	"Hartslag" vs. "heartrate"
Different units	"Mmol/l" vs. "g/l"
Different sources	Blood pressure: femoral, radial
Numeric vs. string	5 vs. ">2"
Granularity	"20-30y" vs. "20", "21"...
Different distributions	Different populations

Data Harmonisation

Distribution   Language

Concept 'PH'

MIMIC

Medical Ontologies

SNOMED CT
ICD10 / CIM10
MedDRA
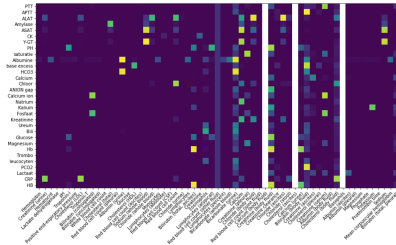Rxnorm / ATC
Clinical Terms Version 3

AmsterdamUMCdb

Matching based on names (string matching)
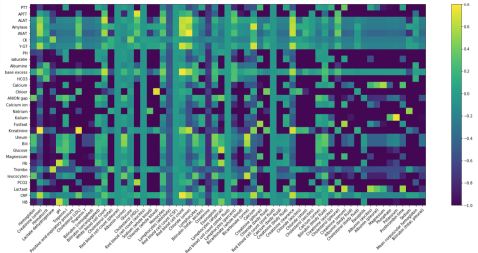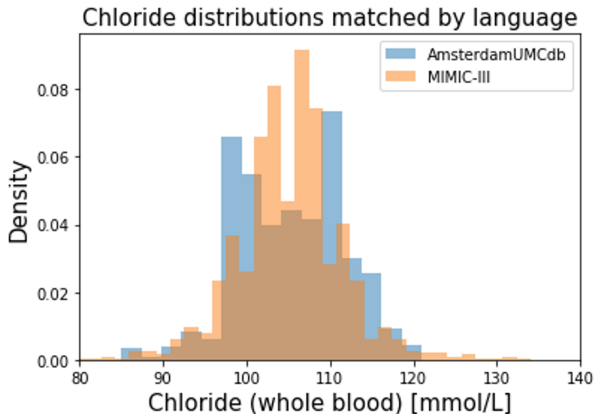
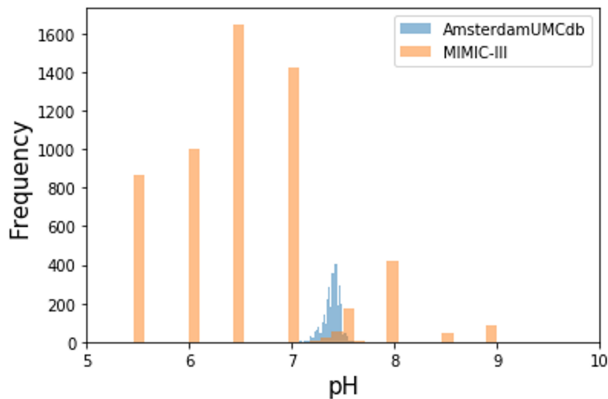# Matching based on data distribution

t-test

Normalised interquartile range overlap

Examples

Examples

Validation Tool

```
Is this a suitable match:  Kreatinine, Creatinine (serum)
Confirm? [Y/N] or X to exit: yes

Is this a suitable match:  Chloor, Chloride (urine)
Confirm? [Y/N] or X to exit: n

Is this a suitable match:  Kreatinine, Creatinine (pleural)
Confirm? [Y/N] or X to exit: n

Is this a suitable match:  Kreatinine, Creatinine (ascites)
Confirm? [Y/N] or X to exit: no

Is this a suitable match:  Albumine, Albumin (urine)
Confirm? [Y/N] or X to exit: exit
```

```
[28] print('Saved Matches:')
     print(matches)
     print('Non_Matches:')
     print(non_matches)
```

```
Saved Matches:
[['PH', 'pH'], ['Calcium', 'Calcium'], ['Natrium', 'Sodium'], ['Kalium', 'Potassium'], ['Fosfaat', 'Phosphate'], ['Kreat
Non Matches:
[['Calcium ion', 'Calcium (urine)'], ['Calcium ion', 'Calcium'], ['Kreatinine', 'Creatinine (urine)'], ['Chloor', 'Chlor
```

Thank you!

Code Link:

https://github.com/EmmaRocheteau/ICUnity