

# DISAMBIGUATION OF ICPC CODES

Tom Seinen

PhD Student – Erasmus MC



This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

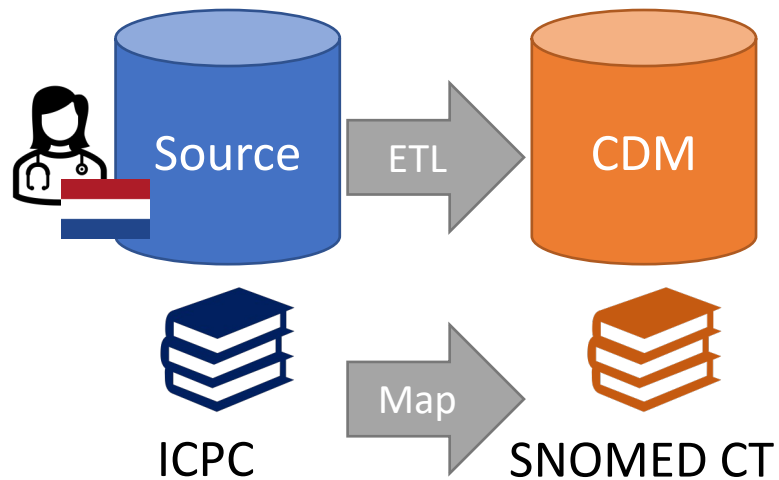


Erasmus MC  
University Medical Center Rotterdam



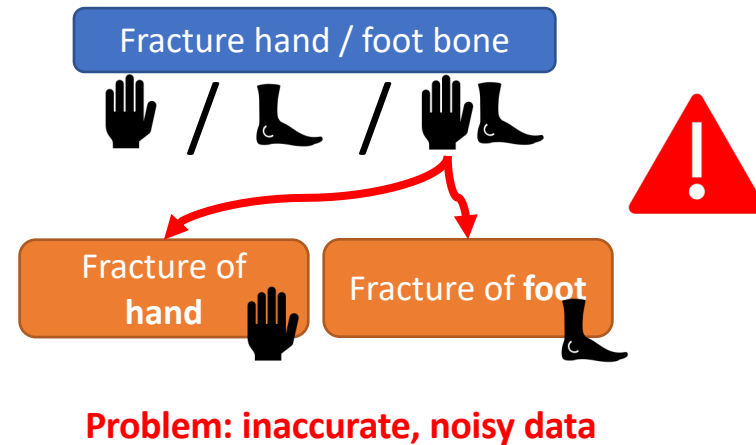
# Context & Problem

- Dutch general practitioner database:
- 2.5 million patients
- International Classification of Primary Care (ICPC)



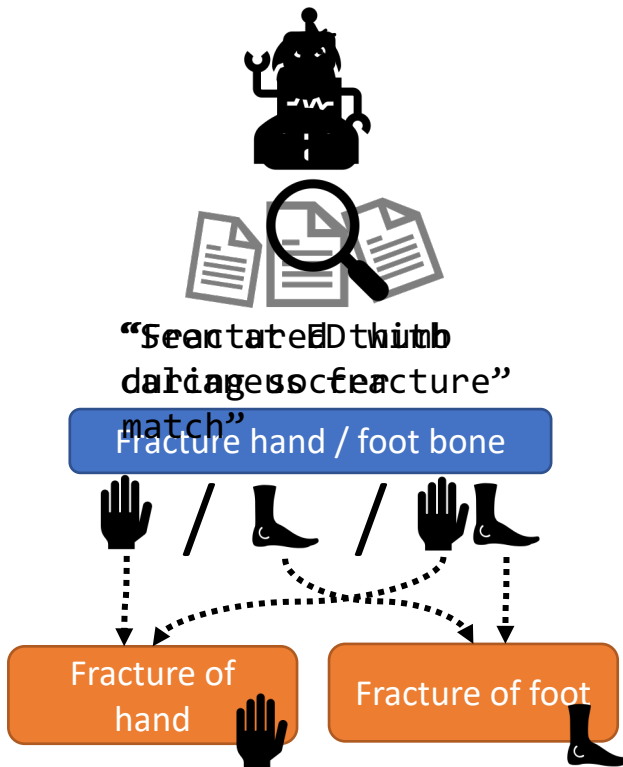
## Difficult to map codes:

- L74: Fracture hand / foot bone
- K78: Atrial fibrillation / flutter
- D75: Malignant neoplasm colon / rectum



# Research question

- **Map single codes manually:**
- Consulting **free-text** clinical notes
- **Around** the registered ICPC code
- **This takes a lot of work:**
- **Thousands** of observations per ICPC code

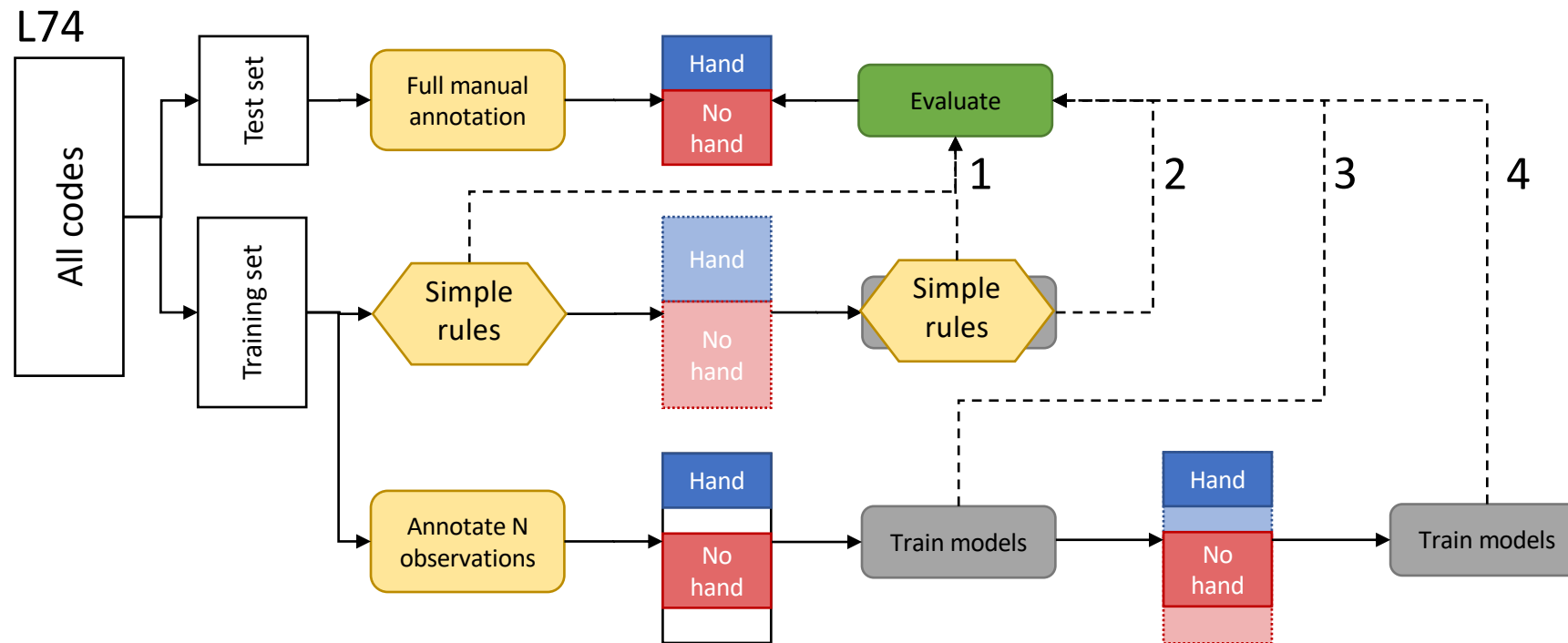


Can we do this automatically?

- **Classification** problem
- **Predict** the right **concepts**
- Given the **surrounding data**

# Methods

- Classifier: Lasso
- Features: TFIDF



Simple rules:  
"hand", "wrist", "finger"  
"foot", "ankle", "toe"

# Results

L74: Fracture hand / foot bone

Performance on test set		Hand		Foot	
	# ann.	AUC	AUPRC	AUC	AUPRC
1. Rules	NA	0.87	0.82	0.90	0.83
2. Rules gen.	NA	0.91	0.91	0.96	0.95
3. Ann.	100	0.97	0.97	0.97	0.96
4. Ann gen.	100	0.97	0.97	0.96	0.94
3. Ann.	200	<b>0.98</b>	0.98	<b>0.99</b>	<b>0.99</b>
4. Ann gen.	200	0.98	0.98	0.99	0.99
3. Ann.	300	0.98	0.98	0.99	0.99
4. Ann gen.	300	0.98	<b>0.99</b>	0.99	0.99

# Results

D75: Malignant neoplasm colon / rectum

Performance on test set		Colon		Rectum	
	# ann.	AUC	AUPRC	AUC	AUPRC
1. Rules	NA	0.76	0.71	0.81	0.62
2. Rules gen.	NA	0.70	0.74	0.84	0.74
3. Ann.	100	0.72	0.70	0.83	0.68
4. Ann gen.	100	0.75	0.73	0.82	0.66
3. Ann.	200	0.81	0.84	0.85	0.77
4. Ann gen.	200	0.83	0.85	0.87	0.80
3. Ann.	300	<b>0.85</b>	<b>0.87</b>	<b>0.90</b>	<b>0.84</b>
4. Ann gen.	300	0.85	0.87	0.89	0.83

# Annotation tool

Lightweight R-shiny application

Runs directly on the CDM

Active learning to sort observations

Visualize feature importance

The screenshot displays the Annotation tool interface, which is a lightweight R-shiny application. The interface is organized into several panels:

- Code occurrence:** 2021-06-18, id: 3, Subject id: 2116, Initial row id: 3.
- Annotation info:** Annotated occurrences: 301, Since last training: 0.
- Note text data:** A list of notes with highlighted words. The first note (849900) describes a fall in a home, mentioning "gevallen", "pijn", "voet", "kan", "met", and "veel". The second note (849901) mentions "hematoom" and "Drukpijn". The third note (849902) mentions "naar", "gebeld", "dat", "ze", "konden", "komen", "Mw", "gaat", "rijden", "aldaar", "rolstoel", "Uitleg", "indien", "naar", "huis", "dan", "geen", "breuk", "anders", "naar", "SEH", "gestuurd", "Afspraak", "diagnostiek", "gemaakt", "bij", "Beeldvormend", "onderzoek", "Gelre", "#City#", "locatie", "#City#" and "onder verwijsnummer: #LongNumber# Reden:". The right side of this panel shows "After code:" with a list of codes and their descriptions, such as "849905: 2021-06-21 CHI (#City#) SEH. Fractuur basis MT5 R. Drukverband met gipschoen. Co gipspoli 1 week." and "849907: 2021-06-28 OOG (#Name#) Cataract operatie RE. Zonder complicaties."
- Structured data:** A panel with a plus sign (+).
- Set annotation:** A panel showing the current annotation status as "Manually annotated". It includes three buttons: "fractuur" (checked), "hand" (unchecked), and "voet" (checked). Below these are "Annotate & Next" (green), "Previous", "Next", "3" (input field), and "Go to id" buttons.
- Prediction result:** A panel showing the probability and class for each annotation: "fractuur: 1", "hand: 0", "voet: 1".
- Highlight:** A panel with a plus sign (+).
- Train model:** A panel with a plus sign (+).
- Sorting method:** A panel with a plus sign (+).

# Conclusion

- **Possible:** Automatically **classify** individual **ICPC codes** into **narrower OMOP concepts**
- **Practical use:** improve concept **mappings**
- **Performance** difference between:
  - ICPC codes
  - Modelling methods
- **More info & future steps** or **questions?**
- Meet me at my **poster**



We classified **ambiguous source codes** into narrower **OMOP concepts**