Tampere University

AI-Driven NLP Models to Identify Aging-Related Health Issues in Free-Text EHR Data

2nd ENFIELD Webinar
Bias in Medical AI:
Identifying Risks and
Ensuring Fairness

23 May 2025

Juulia Jylhävä, Docent, PhD

Senior Research Fellow
Health Sciences and Gerontology
Research Center
Tampere University | Tampere
Finland
juulia.jylhava@tuni.fi

Principal researcher
Department of Medical Epidemiology
and Biostatistics
Karolinska Institutet | Stockholm
Sweden
juulia.jylhava@ki.se

Background

- Many clinically significant health conditions are frequently underreported in diagnostic codes or reported only in textual electronic health records (EHRs)
- Systematic underreporting particularly prevalent in age-related syndromes compared to other conditions (Ugboma et al. CME J Geriatr Med. 2008)
- Many conditions, such as mobility limitations do not have diagnostic codes

  - age-associated conditions carry relevant information regarding treatment and prognosis
- Approximately 80% of medical records are unstructured textual data

  - require minimal maintenance yet contain a wealth of detailed information not captured in structured data, such as diagnostic codes

  - traditional methods, such as manual abstraction not adequate for processing large amounts of textual EHR data

Aim

1. To analyze whether an AI-guided deep learning-based natural language processing (NLP) model can effectively identify commonly underreported age-related health conditions in textual EHRs:

1) incontinence, 2) falls, 3) mobility limitations and 4) loneliness

2. To assess whether the NLP approach would identify a greater number of falls and incontinence cases compared to ICD-10 codes and lead to improved risk stratification of all-cause mortality

Data

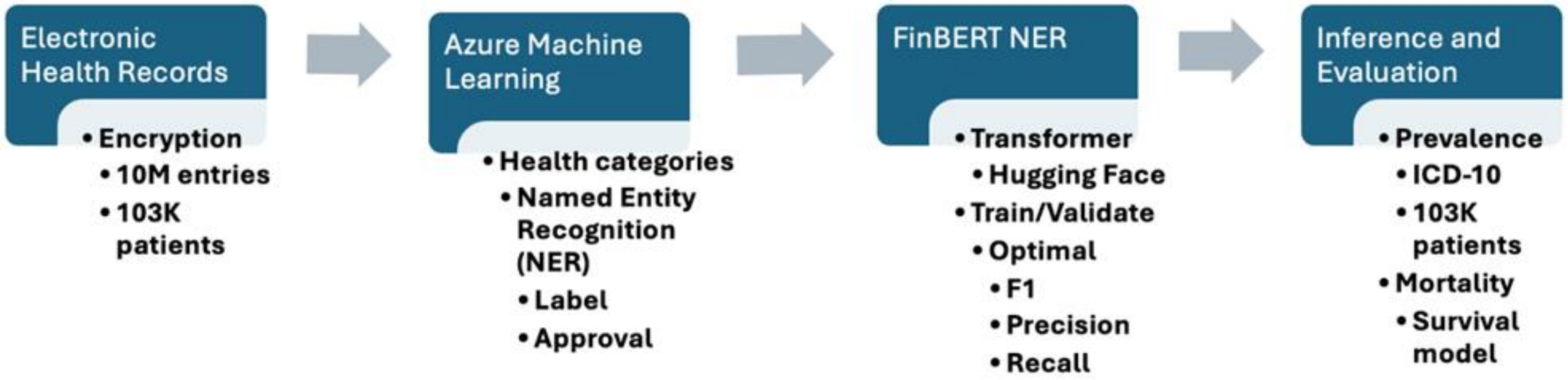
- Patient data from Central Finland wellbeing services county
  - 10.6 million free-text records from 103k individuals
  - Covering primary, secondary, tertiary and institutional care from 2010 to 2022
  - Diagnoses (ICD-10), routine lab tests, outcome data (e.g. mortality)

    - falls W00-W19 (excluding jumping into water/pools W16.5-W16.9)

    - incontinence N39.4, R15, R32 and R39.81
  - Patients aged from 50 to 80 years
  - 52,642 (51.35%) were female
  - 21,213 (20.69%) patients died during the period

Method

- NLP applied with named entity recognition (NER) and Google's Bidirectional Encoder Representations from Transformers (BERT) model for Finnish language (FinBERT): **falls, incontinence, loneliness and mobility limitations**
- Model trained with manual labelling and BIO tagging scheme, assigning each health condition to categories (yes, no) or scale (0, 0.5, 1) for mobility limitations
- NER model evaluation with precision, recall and F1 score against a human rater
- Prevalence estimates by age category and comparison against ICD-10 codes (for falls and incontinence) to assess case identification reliability
- Cox regression models to assess association with all-cause mortality



NER model evaluation

Precision

Positive predictive value

$$\frac{\text{True Positives}}{\text{Predicted Positives}} \quad \text{or} \quad \frac{TP}{TP + FP}$$

Recall

Sensitivity, true positive rate

$$\frac{\text{True Positives}}{\text{Actual Positives}} \quad \text{or} \quad \frac{TP}{TP + FN}$$

>0.7 good
>0.8 very good

$$\text{F1 Score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Examples of labeled text passages

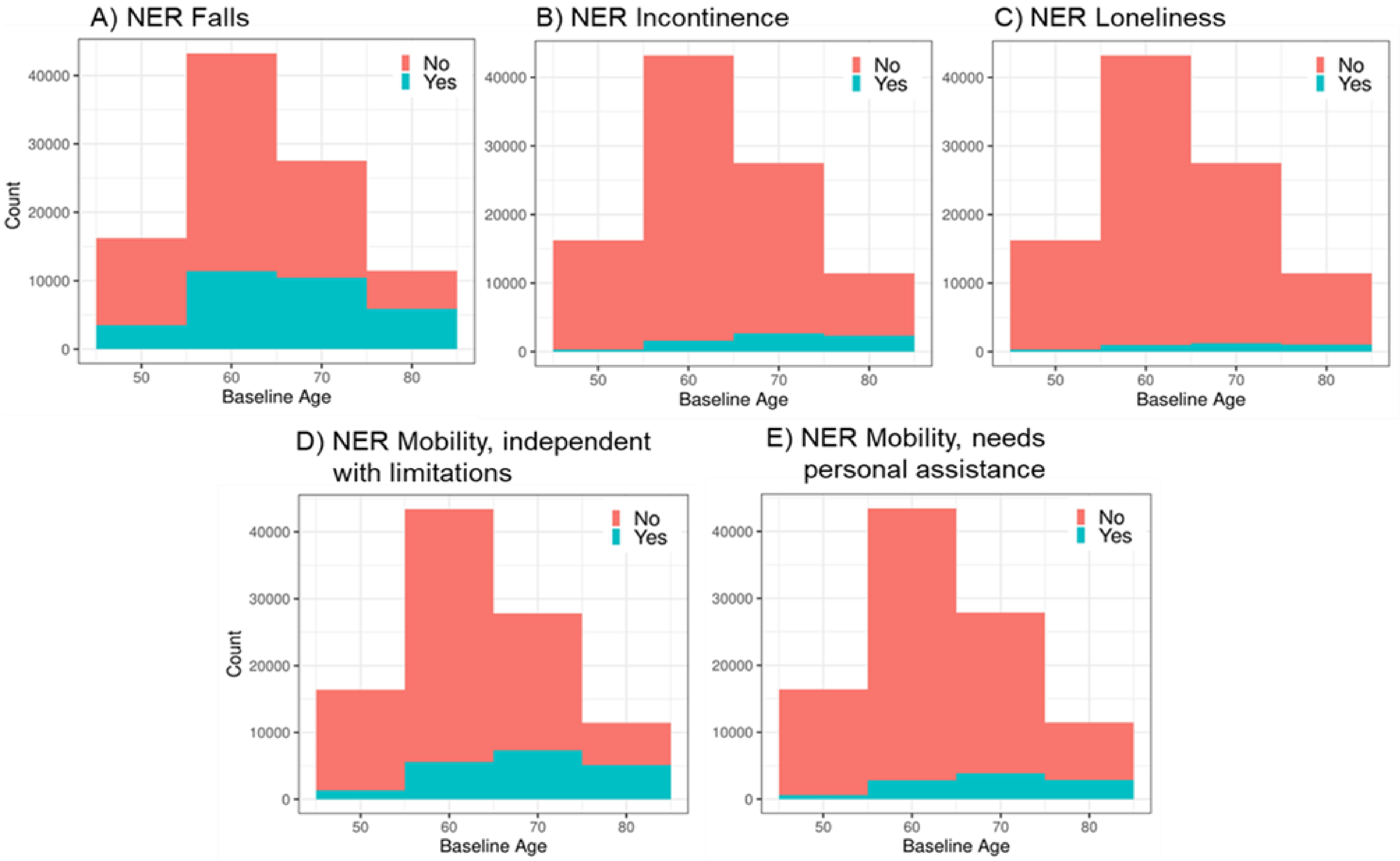
Health condition	Text sequences
Falls	‘patient had a **fall**’, ‘patient **fell**’, ‘patient found **fallen** on the floor, ‘patient **tripped** on to something’, ‘patient has had multiple **falls**’
Incontinence	‘patient is experiencing **loss of bladder/bowel control**’, ‘patient is suffering from **urinary/fecal incontinence**’, ‘patient experiencing **urinary/fecal leakage**’, ‘patient is **unable to withhold urine/stool**’, patient is **unable to control urination/defecation**’
Loneliness	‘patient reports feeling **lonely**’, ‘patient reports feelings of **loneliness**’, ‘patient suffers from **loneliness**’, ‘patient is experiencing **loneliness**’
Mobility limitations, independent with limitations	‘patient is **limping**’ ‘**moving triggers pain**’, ‘patient uses **walking aids [cane/walker/crutches/rollator]**, ‘patient needs to use **walls/handrails for support when walking**’,
Mobility limitations, needs personal assistance or e.g. wheelchair	‘patient needs **personal assistance in ambulation**’, patient can only **move/walk with assistance by someone’**, ‘patient needs **a walker**, ‘patient uses/needs a **wheelchair**’, ‘patient needs **transfer aids**’, ‘patient is **bedbound/non-ambulatory/immobile in bed**’

Results: model evaluation

Health Category	Labeling Sample Size	F1	Recall	Precision	Inferred cases by NER	Cases by ICD-10	NER ICD Intersection
Falls	3,000	0.87	0.86	0.88	31,987	4,090	3,639
Incontinence	3,000	0.81	0.84	0.78	7,059	3,873	1,200
Loneliness	1,000	0.87	0.91	0.84	3,623	NA	NA
Mobility	5,000	0.85	0.86	0.84	29,695	NA	NA

- All F1 scores in the ’very good’ range
- NER models identify a greater number of cases compared to ICD-10 codes
- Overlap not perfect

Proportions by age group



Tampere University

Karolinska Institutet

GEREC GERONTOLOGY RESEARCH CENTER

Results: all-cause mortality

- Using the NER-identified vs ICD-10 -based incontinence and fall cases
- NER models more predictive
- Loneliness not significant
- Mobility limitations predictive with a dose-responsive pattern

	NER-based model			ICD-10-based model		
Health condition	HR	95% CI	p	HR	95% CI	p
Falls						
Full sample	1.31	1.27-1.35	p<2e-16	1.04	0.99-1.10	0.14
Men	1.39	1.34-1.44	p<2e-16	1.12	1.03-1.21	0.006
Women	1.20	1.15-1.25	p<2e-16	0.98	0.90-1.05	0.52
Incontinence						
Full sample	1.99	1.92-2.06	p<2e-16	0.65	0.61-0.70	p<2e-16
Men	2.02	1.92-2.12	p<2e-16	0.61	0.56-0.67	p<2e-16
Women	1.95	1.85-2.05	p<2e-16	0.74	0.66-0.92	0.09
Loneliness						
Full sample	0.97	0.92-1.03	0.37	-	-	-
Men	0.95	0.87-1.04	0.30	-	-	-
Women	0.98	0.91-1.06	0.66	-	-	-
Mobility, independent with limitations				-	-	
Full sample	1.89	1.84-1.96	p<2e-16			-
Men	1.96	1.89-2.05	p<2e-16	-	-	-
Women	1.80	1.72-1.88	p<2e-16	-	-	-
Mobility, needs personal assistance				-	-	
Full sample	2.19	2.12-2.26	p<2e-16			-
Men	2.24	2.15-2.34	p<2e-16	-	-	-
Women	2.12	2.02-2.22	p<2e-16	-	-	-

CI, confidence interval; HR, hazard ratio; ICD-10, International Classification of Diseases, 10th Revision

Risks and challenges

1. Temporal context loss

- the model is not able to discriminate between past and present conditions/symptoms: “patient had a fall two years ago”
- in our approach this was not a major issue; in vast majority of the texts the fall was very recent

2. Attribution error/misattribution

- information that refers to someone other than the patient, such as a spouse, child, or parent is incorrectly assigned to the patient themselves
- present in our data, affecting a very small number of cases

Risks and challenges

3. Negation detection

- NER noun-based, issues with words like “no”
- pertinent in some of our cases where the trunk of the word(s) was similar in positive and negative contexts
- custom code to single out false positives from true positives

4. Use of abbreviations and shortened words

- the model was trained to detect unambiguous abbreviations and shortened words in a correct context
- ambiguous abbreviations were not included in labeling, potentially causing false negatives (to avoid false positives)

Thank you!

Jake Lin
Mikaela von Bonsdorff
Anna Tirkkonen
Anna Kuukka
Juho Kaijansinkko
Antti Kariluoto
Tomi Korpi

Maija Satamo
Markus Haapanen
Hanna Pajulammi
Parinaz Poursafa
Veracell Oy
Central Finland wellbeing services county

Funding

Research Council of Finland

Sigrid Juselius Stiftelse

eurostars™

EUREKA

European Commission

Instrumentarium Science Foundation

Health Data Science profiling area, Faculty of Medicine and Health Technology, Tampere University

Signe & Ane Gyllenberg Foundation