Natural Language Processing to Identify a Cohort of Suicidal Adolescents for an Epidemiological Study of Risk Factors for Self-harm

Sophie Epstein[1,2], Sumithra Velupillai[1,3,4], Andre Bittar[1,3], Catherine Polling[1,3], Rina Dutta[1,3], Johnny Downs[1,3]

[1]NIHR Maudsley Biomedical Research Centre [2]South London and Maudsley NHS Foundation Trust [3]King's College London [4]KTH, Stockholm, Sweden

National Institute for Health Research

Background

- Self-harm occurs in approximately 1 in 10 adolescents
- Important risk factor for subsequent completed suicide
- History of suicidal thoughts increases risk of self-harm
- Information regarding suicidality often recorded in free-text electronic health records (EHRs) rather than structured data

Aim

To use Natural Language Processing (NLP) to extract data from free-text EHRs to identify a cohort of adolescents with suicidal thoughts/behaviours

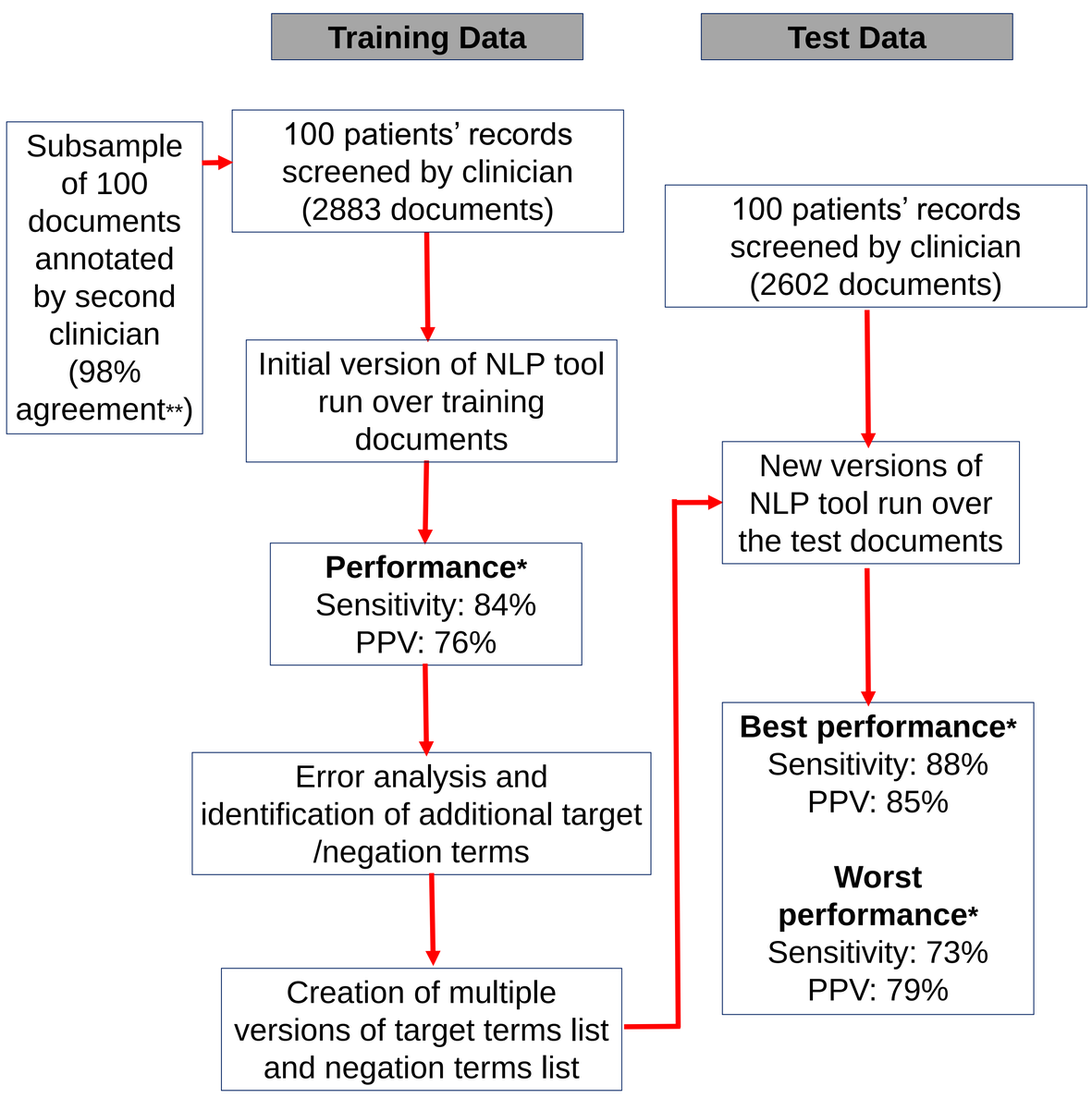
Population and Setting

South London and Maudsley NHS Foundation Trust (SLaM)
11-17 year olds who accessed CAMHS services between 1st April 2009 and 31st March 2013

Data for Natural Language Processing (NLP)
Anonymised EHR dataset of structured data and free text fields for whole sample, obtained using the Clinical Records Interactive Search (CRIS) system

Natural Language Processing

Rule-based NLP tool developed using pyConText[1] software which identifies positive and negated mentions of suicidality



* Patient level performance **% accuracy

1. B.E. Chapman, S. Lee, H.P. Kang, and W.W. Chapman. 2011. Document-level classification of CT pulmonary angiography reports based on an extension of the context algorithm. Journal of Biomedical Informatics, 44(5):728–737

Process for Identifying Suicidality

1. Target terms

Target terms for suicidality:
Suicid*
Kill himself/herself/myself
End his/her/my life
Wish he/she was dead

Example:
'His mother reported that last week he threatened to kill himself'

2. Negation detection

Negation terms:	Experiencer terms:
Not	Mother
Denied	Family
Never	Friend

Example:
'She denies any history of suicidal thoughts'

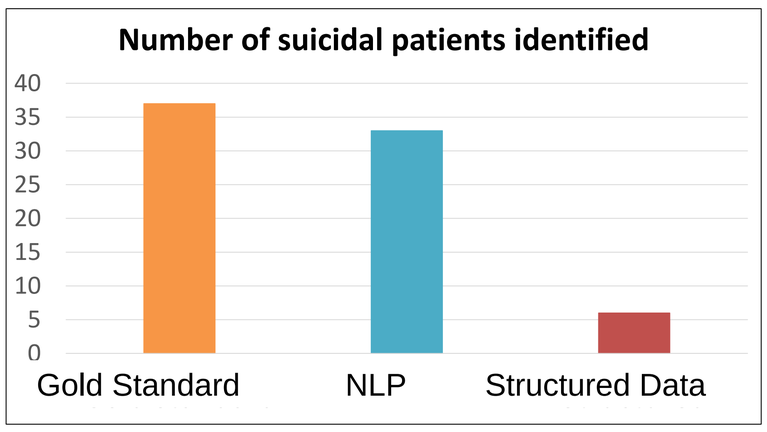
Example:
'Maternal grandmother attempted suicide'

3. Decision rule

Presence of one positive mention of suicidality → patient suicidal (Regardless of presence of other negated mentions)

Results: Comparison of NLP with Structured Data

- 100 patients
- Risk assessment questions on suicide risk
  - Binary: (yes/no)
  - Categorical ('low, low/mod, moderate, mod/high, high'). Cut off of 'mod' was considered positive.



Conclusion and Plans for Further Analysis

The NLP tool was effective in identifying suicidal CAMHS patients from anonymised EHRs.
Planned cohort study using linked data:

*data linkage

CLINICAL RECORD INTERACTIVE SEARCH

MQ Transforming mental health

NHS South London and Maudsley NHS Foundation Trust