Using free text electronic health records from Sussex mental health services to implement a risk calculator to identify people at risk of psychosis

Elizabeth Ford[1], James Stone[1], Benjamin Fell[2], Gergely Bartl[3], Dominic Oliver[4], Paolo Fusar-Poli[4] and Kathryn Greenwood[3,5]

1) Brighton and Sussex Medical School, Brighton, UK; 2) Akrivia Health, Oxford, UK; 3) University of Sussex, Brighton, UK; 4) Institute of Psychiatry, Psychology and Neuroscience, King's College London; 5) Sussex Partnership NHS Foundation Trust

NIHR | Applied Research Collaboration Kent, Surrey and Sussex

Introduction

There are high rates of emotional disorders and substance use among young people in Sussex.

These experiences are risk factors for **Psychosis.**

95% of **at-risk** young people are not recognized as at risk even when seen by secondary mental health services.

Electronic health records

Mental health consultations in Sussex are captured in written notes in an electronic patient record system. These capture lots of details about patients' experiences and symptoms.

These are now processed via Akrivia Health Ltd in the Clinical Record Interactive Search system (CRIS).

Data from clinic notes can be extracted, structured and used to build risk stratification tools.

A risk calculator tool predicting **first episode psychosis** was recently developed and validated in London.

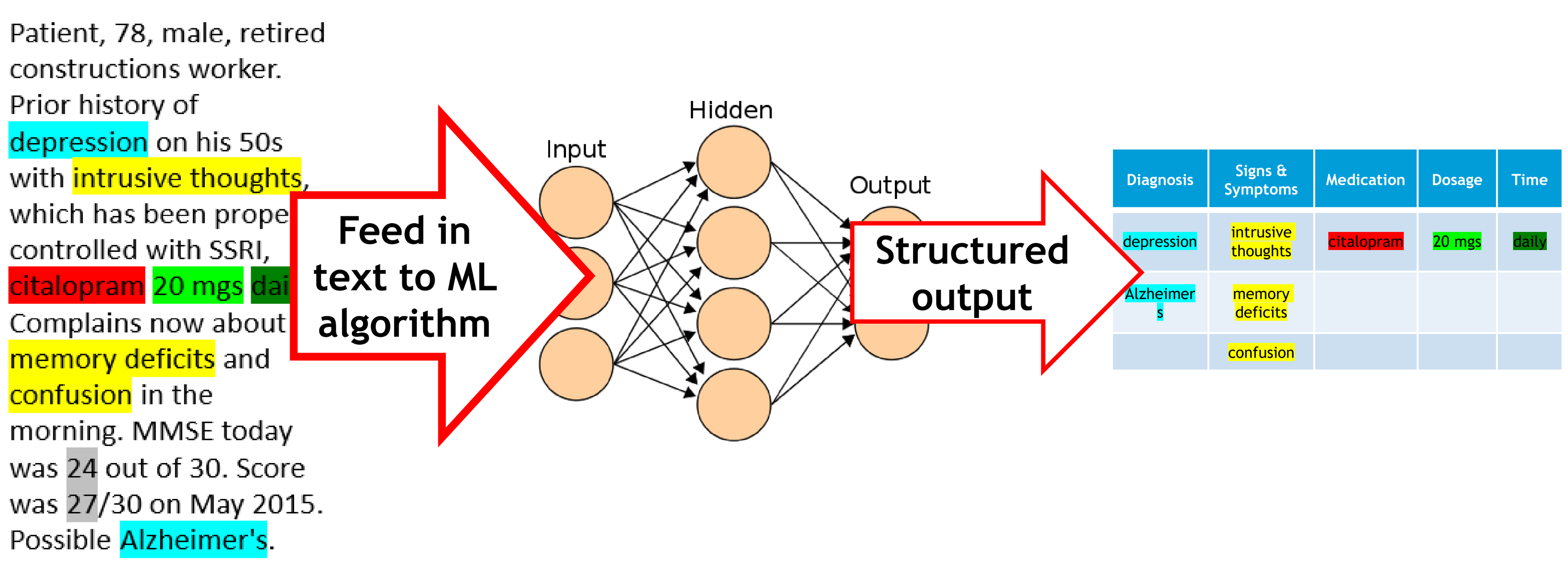
Objectives

We aimed to recreate and validate the London risk calculator for psychosis within Sussex NHS Mental Health trust data. We also consulted with patients about how it should be implemented and what support patients should be given if found to be at risk.

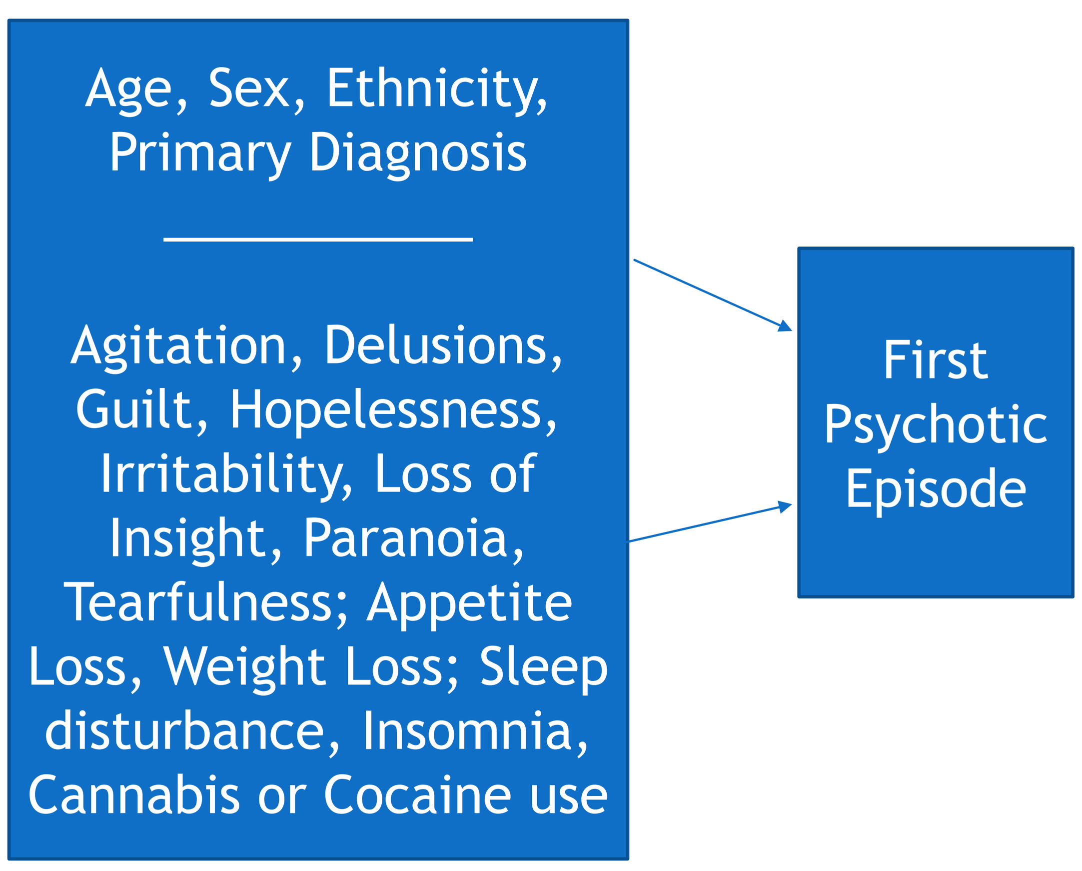
Methods

We annotated mental health clinic notes data for 14 concepts: Agitation, Delusions, Guilt, Hopelessness, Irritability, Loss of Insight, Paranoia, Tearfulness; Appetite Loss, Weight Loss; Sleep disturbance, Insomnia, Cannabis and Cocaine use.

The annotations were used to train a machine learning algorithm to extract the concepts from unseen patient notes



Structured data was then fed into a Cox Regression model to predict the outcome (time to psychosis diagnosis). The model is evaluated for accuracy using Harrell's Concordance Statistic (Harrell's C).



Results

The natural language processing (NLP) algorithms were found to be accurate at extracting the clinical information

Irving Concept	Training Samples	Validation Samples	Precision	Recall	F1
Hopelessness	72	21	90%	86%	88%
Poor insight	106	22	64%	82%	72%
Tearfulness	205	59	97%	100%	98%
Irritability	116	33	91%	94%	93%
Agitation	305	65	91%	91%	91%
Guilt	79	31	90%	90%	90%
Paranoia	217	50	80%	83%	81%
Delusions	214	46	84%	89%	86%
Appetite (loss)	302	70	92%	92%	92%
Weight (loss)	175	49	91%	82%	86%
[Sleep Quality - Good]	381	85	78%	88%	83%
[Sleep Quality - Poor]	1412	351	73%	88%	80%
Substance us - cocaine	78	24	93%	89%	91%
Substance use - cannabis	226	52	94%	91%	92%

F1 Score is an estimation of accuracy: perfect would be 100%

There were 144,916 Sussex patients with suitable data, including a baseline predictor diagnosis, sex, age and ethnicity; and 13,295 (9.2%) of these were subsequently diagnosed with psychosis.

The preliminary Cox regression model, using all the predictors, predicted the outcome of psychosis with Harrell's C of 0.655.

Next Steps

The accuracy of the model so far is not good enough for clinical implementation. We are exploring several avenues to improve the model in Sussex:

- Demographics in Sussex are different from South London so weighting for demographic variables might need to be changed.
- The London calculator used coded data for diagnosis for both predictor and outcome. Only 4% of Sussex patients had a coded diagnosis so we relied on NLP for identifying diagnoses. There is substantial room for improving the NLP extraction of diagnosis which we are working on.
- Only clinic notes, and not letters and reports, were available for information extraction. Letters and reports are more likely to have diagnostic information. We will incorporate these into the NLP data extraction to get richer information on patients.
- We could consider additional predictors or a slightly different outcome for an adapted model to try to get higher accuracy.

Conclusions

Replicating a risk calculator in a new NHS Trust's patient data is not straightforward.

Data types, population demographic and clinical structures are different and pose challenges for the portability of risk calculators, suggesting they may require adaptation and further validation.

Contact

Elizabeth Ford @ BSMS e.m.ford@bsms.ac.uk

Benjamin Fell @ Akrivia benjamin.fell@akriviahealth.com

This work is supported by the NIHR Applied Research Collaboration in Kent Surrey and Sussex (grant no. NIHR200179). Views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

brighton and sussex medical school

AKRIVIA HEALTH

INSTITUTE OF PSYCHIATRY, PSYCHOLOGY & NEUROSCIENCE

UNIVERSITY OF SUSSEX