Enriching psychotic disorder classification using natural language processing

Rashmi Patel, Richard Jackson, Robert Stewart, Philip McGuire

NHS
National Institute for
Health Research

Background

The classification of psychotic disorders is based on criterion-based diagnostic systems (such as ICD-10 and DSM-5) which do not necessarily reflect their underlying aetiology and pathophysiology. A more refined characterisation of clinical phenotype could help to improve our understanding of these disorders. Clinical data are increasingly recorded in the form of electronic health records (EHRs). Natural language processing (NLP) offers the opportunity to quickly analyse large volumes of detailed clinical data from EHRs. We sought to characterise the breadth of presenting symptoms in people with psychotic disorders using NLP.

Methods

Dataset: South London and Maudsley NHS Trust (SLaM) Biomedical Research Centre (BRC) Case Register comprising pseudonymised EHRs of over 270,000 people.

Data collection: The NLP software package TextHunter was used. All sentences containing keywords relevant to psychotic symptoms (Table 1) were analysed using a support vector machine learning (SVM) approach.

Outcomes: prevalence of psychotic symptoms and their association with ICD-10 diagnosis.

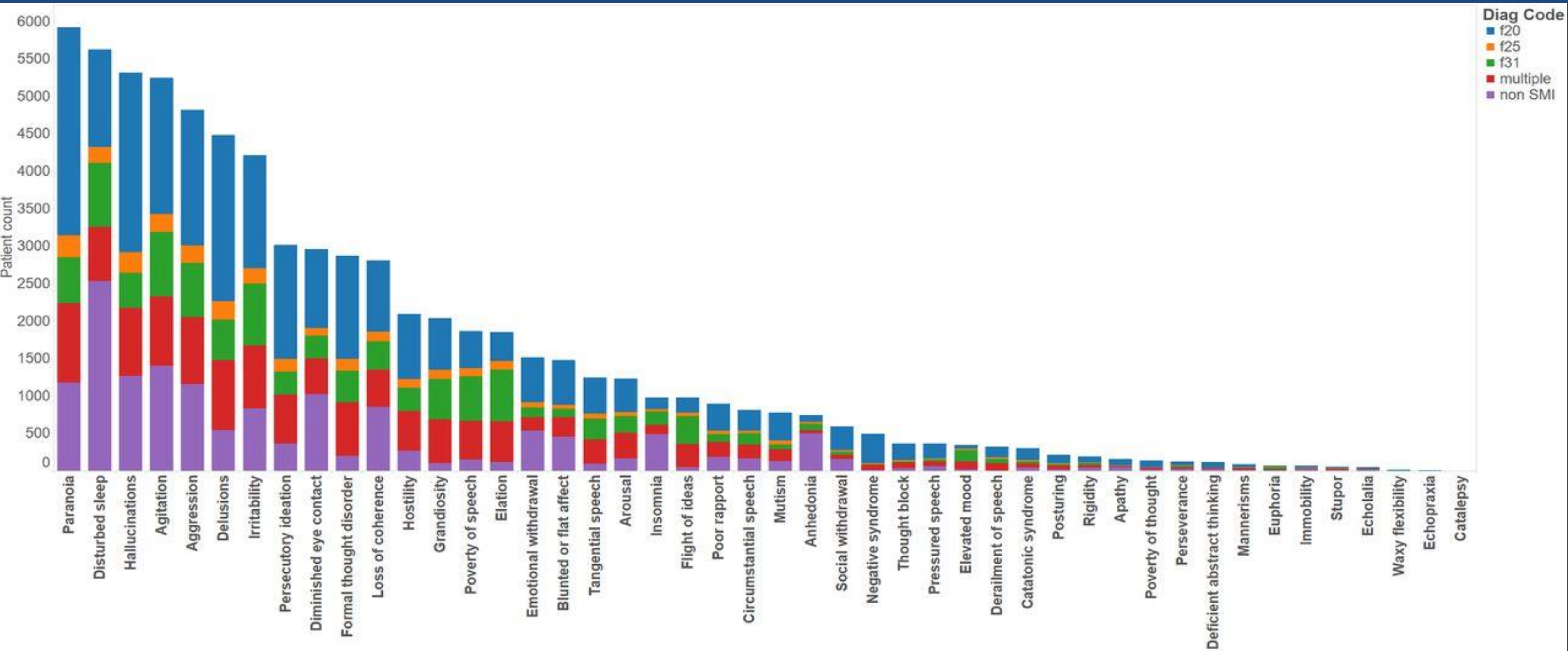## Figure 1 – Distribution of psychotic symptoms across diagnoses

Table 1 – Psychotic symptom domains

Domain	Symptoms
Positive	Agitation, aggression, arousal, hostility, delusions, hallucinations, paranoia, persecution
Negative	Diminished eye contact, blunted or flat affect, emotional withdrawal, social withdrawal, abstract thinking, poor rapport, apathy, anhedonia, poverty of speech, poverty of thought, negative syndrome
Disorganised	Circumstantial speech, reduced coherence, formal thought disorder, thought block, tangential speech, derailment, flight of ideas
Manic	Elevated mood, disturbed sleep, insomnia, euphoria, pressured speech, irritability, elation, grandiosity
Catatonic	Mannerism, rigidity, posturing, perseverance, stupor, waxy flexibility, immobility, echolalia, mutism, catalepsy, echopraxia

Results

Data on 46 psychotic symptoms were obtained from 18,761 patients with an ICD-10 diagnosis of a psychotic disorder (F20, F25 or F31) and a control group of 57,999 patients without a psychotic disorder diagnosis. The most frequently documented symptoms were paranoia, disturbed sleep and hallucinations (Figure 1). Psychotic symptoms were not limited to patients with an ICD-10 diagnosis of a psychotic disorder and were also present in the control group. However, the distribution of symptoms varied depending on diagnosis with disturbed sleep and anhedonia being frequently documented in the control group and paranoia, hallucinations and delusions being more frequently documented in patients with a schizophrenia diagnosis.

Discussion

We obtained detailed symptom data from a large sample of patients using NLP. We found that psychotic symptoms were not limited to patients with a specific ICD-10 diagnosis and were present in a wide range of ICD-10 disorders. These findings highlight the utility of detailed NLP-derived symptom data to better characterise psychotic disorders and improve our understanding of their underlying pathophysiology.

NHS
South London
and Maudsley
NHS Foundation Trust