

The application of Natural Language Processing to real-world electronic health record data to identify positive and negative symptoms in schizophrenia

Rashmi Patel^{1,2}, Rifqi Alkhatib², Yida Won², Soon Nan Wee², Simran Thadani², Lu Wang² and Miguel E. Rentería²

1. King's College London, London, UK. 2. Holmusk Technologies, Inc., New York, NY, USA.



BACKGROUND

- Schizophrenia is a heterogenous disorder, presenting with positive, negative and cognitive symptoms. It is associated with significant disability and impairment across all aspects of life.¹
- Response to treatment is variable, with psychotropic medication more effective for positive symptoms.²
- The use of electronic health record (EHR) data combined with techniques such as Natural Language Processing (NLP)³ to extract information from unstructured text, has the potential to improve our understanding of the course of mental illnesses, through the generation of real word data.

- Understanding how symptom profiles affect illness severity and hospitalisation may guide treatment focus and enable more preventative steps to minimize disruption to daily life.
- Objective:**
- 1) To establish if NLP can identify the presence of positive and negative symptoms in the EHR's of people diagnosed with schizophrenia.
 - 2) To ascertain if positive and negative symptoms are associated with baseline clinical severity, measured using the Clinical Global Impression – Severity Scale (CGI-S), and with psychiatric hospitalisation.

METHOD

Data source

Data from 4,440 patients with schizophrenia receiving care from 25 U.S. mental healthcare providers using the MindLinc EHR system were analysed using Holmusk's NeuroBlu platform (version 21R1).

Methodology

Inclusion criteria: Diagnosis of Schizophrenia (ICD-9 295*; ICD-10: F20*)

Exclusion Criteria: Patients from EHR sites with no inpatient facilities. Patients with no baseline MSE, CGI-S data or antipsychotic prescription data, and less than 1 year of encounter data.

Index Date: The date of the first recorded clinical event in the EHR.

Exposure: NLP was applied to our real-world EHR dataset, where schizophrenia-related symptoms were extracted from all clinician recorded notes of the Mental State Examination (MSE) within a 3-month period from the index date. This resulted in data on 14 different positive and 15 different negative symptoms associated with schizophrenia (Table 1).

Outcome measure: CGI-S at index date; total no. of days spent in a psychiatric hospital within 12 months of the index date.

Analysis

The association between the number of positive and negative symptoms and CGI-S at index date and the number of days spent in a psychiatric hospital within 12 months of the index date were analysed using multiple linear regression with age, gender, race, and marital status as covariates.

NeuroBlu™ database

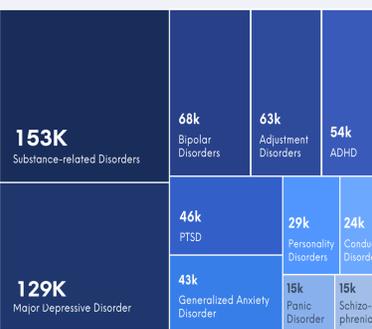


Figure 1. NeuroBlu Database overview

Structured Data

- Outcome Measures (e.g., CGI-S, GAF)
- Diagnosis Codes (ICD-9, ICD-10)
- Prescription Data
- Patient Demographics
- Emergency Department, inpatient & outpatient data across the same patients in 20 of 25 clinics

Unstructured Data

Mental Status Examination (MSE)

- A structured summary of observations of a patient's mental experiences and behaviour at a point in time.
- Holmusk developed >30 advanced Neural Network models to predict structured labels from MSE
- Created >300 psychiatry specific labels in collaboration with clinicians to track disease progression over time

External Stressors

- Social, relational and occupational events that may affect the patient's mental health

Data Source of US Health Facilities

De-identified EHR data were obtained from U.S. mental health services that use the MindLinc EHR system. The data were analysed in NeuroBlu, a secure Trusted Research Environment (TRE) that enables data assembly and analysis using an R/Python code engine.



Figure 2. State specific data source for NeuroBlu

RESULTS

Table 1. Frequency and prevalence of positive & negative MSE symptoms within 3 months of index date

Rank	Positive Symptom	No. of patients	Percentage of Cohort (%)	Negative Symptom	No. of patients	Percentage of Cohort (%)
1	Experiencing delusions/abnormal thoughts (paranoia)	1766	39.8	Affect - blunted/restricted	2236	50.4
2	Experiencing hallucinations (auditory)	1620	36.5	Appearance - Issues with Grooming/Hygiene	1273	28.7
3	Experiencing delusions (not specified)	1287	29.0	Appearance - Issues with Eye Contact	595	13.4
4	Experiencing hallucinations (visual)	1059	23.9	Psychomotor - retarded; slowed	572	12.9
5	Responding to internal stimuli	949	21.4	Speech - impoverished	449	10.1
6	Experiencing delusions/abnormal thoughts (grandeur)	299	6.7	Appearance - Issues with Affect	437	9.8
7	Experiencing delusions/abnormal thoughts (obsessions)	114	2.6	Cognition - issues with attention	410	9.2
8	Experiencing hallucinations (tactile)	73	1.6	Executive Functioning - some impairment	330	7.4
9	Experiencing hallucinations (not specified)	62	1.4	Cognition - issues with concentration	291	6.6
10	History of hallucinations	42	0.9	Speech - mute, non-verbal	95	2.1
11	Experiencing hallucinations (olfactory)	24	0.5	Appearance - Issues with dress	70	1.6
12	Experiencing delusions/abnormal thoughts (persecution)	4	0.1	Psychomotor - catatonic	18	0.4
13	Experiencing delusions/abnormal thoughts (religious)	4	0.1	Intelligence - issues with intelligence	9	0.2
14	Experiencing delusions/abnormal thoughts (sexual)	1	0.0	Language - non-verbal/mute	6	0.1
				Attention/Concentration - issues due to MH	3	0.1

Figure 3. The association between number of positive and negative symptoms, clinical severity (3A) and days of hospitalisation (3B).

Figure 3A. Symptom Burden – CGI-S: CGI-S Score (closest to index date) vs No. of Positive/Negative Symptoms

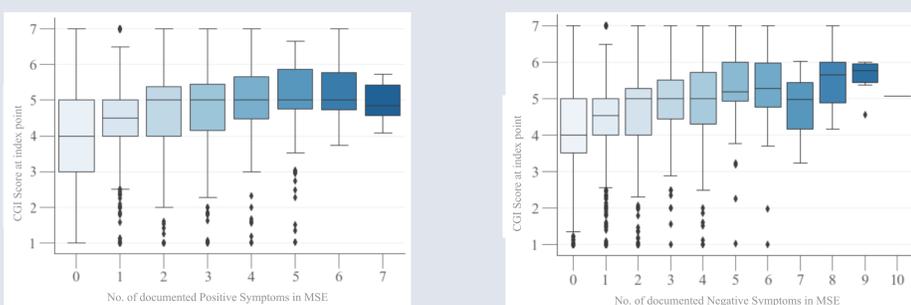
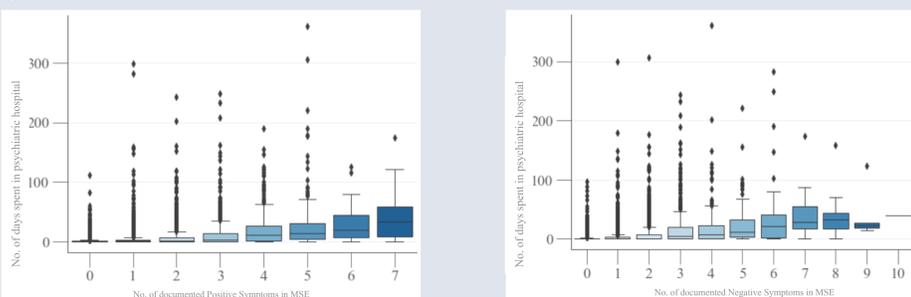


Figure 3B. Symptom Burden- Hospitalisation: No. of Inpatient Days (within 12 months of index date) vs No. of Positive/Negative Symptoms



- 4,440 patients were included in the study, with a mean age of 39.4 years (standard deviation: 15.3). 63% were male, 58.6% from a minority ethnic background and 68.1% single.
- A blunted or restricted affect was the commonest documented symptom, reported in 50% of the cohort (Table 1).
- A higher frequency of both positive symptoms ($\beta = 0.15$, 95% CI 0.13 to 0.18, $p < 0.001$) and negative symptoms ($\beta = 0.11$, 95% CI 0.09 to 0.14, $p < 0.001$) were associated with greater CGI-S at index date (Table 2 & Figure 3).
- A higher frequency of both positive symptoms ($\beta = 2.74$, 95% CI 2.31 to 3.18, $p < 0.001$) and negative symptoms ($\beta = 3.13$, 95% CI 2.70 to 3.57, $p < 0.001$) were associated with a greater number of inpatient days within 12 months of the index date. For every increase in 1 positive symptom there was a 2.74 day increase in length of hospital stay (Table 2 & Figure 3).
- Race (other or Native Hawaiian or other Pacific islander), older age, marital engagement and greater no. of symptoms (positive & negative) were positively associated with the length of hospitalisation, whereas other marital status¹ (divorced, married, separated, death of spouse) were negatively associated with length of hospitalisation (Table 2).

Table 2. Multiple Linear Regression analysis showing covariates that are significantly associated with higher CGI-S scores and a greater number of hospitalisations.

Covariate	CGI-S		Hospitalisation		
	p-value	β Coefficient ¹	p-value	β Coefficient ²	
Gender – Female	0.023	0.09	Race – Native Hawaiian or Other Pacific Islander	<0.001	11.29
Race – Black or African American	<0.001	-0.22	Race – Other Race ³	0.011	5.06
No of Positive Symptoms	<0.001	0.15	Marital Status – Death of Spouse	0.008	-6.73
No of Negative Symptoms	<0.001	0.11	Marital Status – Divorced	<0.001	-4.87
Age	0.002	-0.004	Marital Status – Engaged	0.010	38.15
			Marital Status – Married	<0.001	-5.07
			Marital Status – Separated	0.007	-6.29
			No of Positive Symptoms	<0.001	2.74
			No of Negative Symptoms	<0.001	3.13
			Age	<0.001	0.08

¹ Change in CGI-S score
² Number of psychiatric inpatient days within 12 months
³ Other people of single racial groups that are not specified to be collected by the US Office of Management & Budget (OMB); OR those who are of mixed race

DISCUSSION

- A higher frequency of positive and negative symptoms were associated with greater illness severity and increased number of days hospitalised. The frequency of negative symptoms had a significant impact on the number of days hospitalised in addition to the more well recognised association between hospitalisation and psychotic symptoms⁷.
- Our study aligns with previous literature highlighting the impact of age, and marital state on the risk of hospitalization.^{5, 6}. We also build on growing evidence that women do not experience a milder course of illness than men⁸.
- NLP is an effective tool to help identify data on positive and negative symptom burden from EHRs and could be used to support the development of more targeted treatments.

Limitations

- The MSE symptoms examined were captured during a 3-month window from index date and thus may not capture fluctuations in illness over the 12 months.

References:

1. www.who.int/news-room/fact-sheets/detail/schizophrenia. Accessed 20.10.22
2. Evans LJ, Harris V, Newman L, et al. Rapid and frequent psychiatric readmissions: associated factors. *Int J Psychiatry Clin Pract*. 2017; 21(4):271-276.
3. Correll CU, Schooler NR. Negative Symptoms in Schizophrenia: A Review and Clinical Guide for Recognition, Assessment, and Treatment. *Neuropsychiatry Dis Treat*. 2020 Feb 21;16:519-534.
4. Patel R, Wee SN, Ramaswamy R, et al. NeuroBlu, an electronic health record (EHR) trusted research environment (TRE) to support mental healthcare analytics with real-world data. *BMI Open* 2022;12:e057227.
5. Morgan V, Korsten A, Jablensky A. Modifiable Risk Factors for Hospitalization Among People with Psychosis: Evidence from the National Study of Low Prevalence (Psychotic) Disorders. *Australian & New Zealand Journal of Psychiatry*. 2006;40(8):663-670.
6. Cheng P, Wang L, Xu L, et al. Factors Related to the Length of Stay for Patients With Schizophrenia: A Retrospective Study. *Front Psychiatry*. 2020;12:18254.
7. Mansouri N et al. Risk factors associated with psychiatric hospitalization among Iranian schizophrenic patients. *Iran J Psychiatry Behav Sci*. 2013 Fall;7(2):16-23.
8. Walker S et al. Clinical and social factors associated with increased risk for involuntary psychiatric hospitalisation: a systematic review, meta-analysis, and narrative synthesis. *Lancet Psychiatry*. 2019;6(12):1039 – 1053.
9. Sommer, I.E., Tiihonen, J., van Mourik, A. et al. The clinical course of schizophrenia in women and men—a nation-wide cohort study. *npj Schizophr* 6, 12 (2020)

Conflicts of Interest: All authors report current employment with Holmusk Technologies, Inc. RP reports equity ownership in Holmusk Technologies, Inc.

