However, the clinical classification of psychotic disorders has remained largely unchanged and 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). Automated information extraction methods such as natural language processing (NLP) offer the opportunity to quickly extract and analyse large volumes of clinical data from EHRs. We sought to characterise the range of presenting symptoms in a large sample of patients 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.

Clinical sample: 18,761 patients with an ICD-10 diagnosis of a psychotic disorders (F20, F25 or F31) and a control group of 57,999 patients with a non-psychotic disorder diagnosis (mood/affective/personality disorders without psychotic symptoms).

Data collection: The NLP software package TextHunter was used. All sentences containing keywords relevant to the following symptom categories were analysed using a support vector machine learning (SVM) approach: positive symptoms, negative symptoms, disorganisation, mania and catatonia. Data on 46 symptoms were obtained with 37,211 instances annotated to contribute training and gold standard data for machine learning. 2,950 instances were independently annotated to determine inter-annotator agreement.

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

Results: A good degree of inter-annotator agreement was achieved (Cohen's κ: 0.83). Machine learning NLP achieved a mean precision (positive predictive value) of 83% and recall (sensitivity) of 78%. Among patients with psychotic disorders, the most frequently documented symptoms were paranoia, disturbed sleep and hallucinations. Psychotic symptoms were not limited to patients with an ICD-10 diagnosis of a psychotic disorder and were also present in the control group.

Discussion: 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.

T102. AN INVESTIGATION OF SCHIZOPHRENIA-BIPOLAR SUBGROUPS WITH GENETIC AND PROGNOSTIC VALIDATION

Dominic Dwyer*[,1], Janos Kalman[1], Monika Budde[1], Urs Heilbronner[1], Anne Ruef[1], Heike Anderson-Schmidt[2], Katrin Gade[4], Nikola Mueller[5], Ivan Kondofersky[3], Sergi Papiol[1], Peter Falkai[1], Thomas G. Schulze[2], Nikolaos Koutsouleris[1]
[1]Ludwig Maximilian University; [2]University Medical Centre Göttingen; [3]Institute of Computational Biology

Background: Separation of individuals into schizophrenia and bipolar diagnoses has long been questioned, with some suggesting that the classification impairs the understanding of etiology, the accuracy of prognoses, and treatment selection. In this study, we employed unbiased statistical techniques to identify subgroups of individuals with chronic illness using a large array of variables commonly evaluated at the bedside. We then validated the resulting groups by investigating age of onset, schizophrenia polygenic risk scores (PRS), and functional outcomes at a 1-year follow-up period. Our hypothesis was that transdiagnostic subgroups would be stratified based on illness onset whereby individuals with earlier onset would have higher genetic risk loading and poorer functional outcomes.

Methods: Participants were selected from a longitudinal, naturalistic, multisite project (PsyCourse) designed to investigate psychiatric illness course and outcomes. A total of 329 participants (age(SD)=45.7(12.6); 54% female; years of illness duration(SD) = 13.7(10.3)) with a DSM-IV diagnosis of schizophrenia, schizoaffective disorder, or bipolar disorder were assessed from 17 centers at baseline and 1-year follow-up periods. A clinical battery measuring sociodemographic, illness history, symptoms, cognition, and personality questionnaires (199 variables) was used to subgroup individuals. A non-negative factor analytic consensus clustering MATLAB toolbox was created based on previous methodological work in oncology. PRS were generated using widely used strategies, and differences between resulting subgroups were investigated with MANCOVA controlling for ancestry effects. Differences in functional outcomes were investigated with repeated measures ANOVA.

Results: A 4-subgroup solution was robustly defined as the optimal solution using resampling techniques and cluster validity indices. Diagnoses were mixed in two subgroups, but predominantly bipolar or schizophrenia in the other two. All subgroups had equal illness durations ($p>0.05$), but the age of onset showed a decreasing trend with the earliest age being linked to two subgroups: a mixed bipolar-schizophrenia group with intermediate levels of general functioning and in a schizophrenia group with low levels of functioning ($p<0.001$). PRS scores were significantly increased in the early-onset, mixed bipolar-schizophrenia subgroup ($p=0.007$, uncorrected) and in the schizophrenia group ($p=0.025$, uncorrected). Prognoses differed between the four groups ($p=0.003$), with the greatest increases in functional outcomes in a late-onset mixed diagnostic subgroup ($p=0.006$) and in the schizophrenia group ($p=0.002$).

Discussion: Four subgroups were detected and our hypothesis was supported by a relationship between earlier illness onset and higher schizophrenia genetic risk loading. While one of the subgroups with an earlier onset mostly consisted of individuals with schizophrenia, the other subgroup was diagnostically mixed. Our results tentatively suggest that transdiagnostic clustering may identify subgroups that could be effectively used to understand etiology and prognoses. Future research will investigate the possibility of differential treatment effects in these subgroups.

T103. ODIP (OUTIL DE DIAGNOSTIC INFORMATISÉ DES PSYCHOSES / PSYCHOSIS COMPUTERIZED DIAGNOSTIC TOOL): A NEW, SIMPLE METHOD FOR GENERATING DSM DIAGNOSES FOR PSYCHOTIC DISORDERS

Jean-Romain Richard*[,1], Baptiste Pignon[1], Franck Schurhoff[1], Andrei Szoke[1]
[1]INSERM

Background: OPCRIT was designed as a powerful tool to diagnose psychotic and affective psychoses. It has been frequently used in international psychiatric research. However, with 90 items it is time-consuming to complete and the diagnoses provided include many which are no longer used. Furthermore, this application is no updated for certain operating systems or psychiatric classifications.

For these reasons, we have developed, a similar but much simpler tool focused on DSM classification of affective and non-affective psychoses.

Methods: ODIP is based on the DSM-IV psychotic disorders classification, focusing on psychotic disorders (affective and non-affective). We identified 13 criteria that allow for the distinction between affective disorders with psychotic features (Bipolar or Depressive episode), schizophrenia, schizophreniform, schizoaffective, delusional, brief or non-specified psychotic disorders. We also designed a form to collect data on these 13 items.

To assess how ODIP performs we tested it against the more complete OPCRIT and discordances in diagnosis were compared with the clinical diagnosis or, in a subsample of patients, with a research diagnosis.

This was done in a total sample of 464 patients with a first episode of psychosis.

First, we observed that only 34 out of 90 OPCRIT items are required to obtain a coherent DSM-IV diagnosis and that we could complete the items automatically using an algorithm based on the ODIP form.