


BMJ Open NeuroBlu, an electronic health record (EHR) trusted research environment (TRE) to support mental healthcare analytics with real-world data

Rashmi Patel ,^{1,2} Soon Nan Wee,¹ Rajagopalan Ramaswamy,¹ Simran Thadani,¹ Jesisca Tandi,¹ Ruchir Garg,¹ Nathan Calvanese,¹ Matthew Valko,¹ A John Rush,³ Miguel E Rentería,¹ Joydeep Sarkar,¹ Scott H Kollins^{1,4}

To cite: 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. *BMJ Open* 2022;**12**:e057227. doi:10.1136/bmjopen-2021-057227

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-057227>).

Received 09 September 2021
Accepted 29 March 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Holmusk Technologies Inc, New York, New York, USA

²Department of Psychosis Studies, King's College London, Institute of Psychiatry Psychology and Neuroscience, London, UK

³Curbstone Consultant LLC, Santa Fe, New Mexico, USA

⁴Duke University School of Medicine, Durham, North Carolina, USA

Correspondence to

Dr Rashmi Patel;
rashmi.patel@holmusk.com

ABSTRACT

Purpose NeuroBlu is a real-world data (RWD) repository that contains deidentified electronic health record (EHR) data from US mental healthcare providers operating the MindLinc EHR system. NeuroBlu enables users to perform statistical analysis through a secure web-based interface. Structured data are available for sociodemographic characteristics, mental health service contacts, hospital admissions, International Classification of Diseases ICD-9/ICD-10 diagnosis, prescribed medications, family history of mental disorders, Clinical Global Impression—Severity and Improvement (CGI-S/CGI-I) and Global Assessment of Functioning (GAF). To further enhance the data set, natural language processing (NLP) tools have been applied to obtain mental state examination (MSE) and social/environmental data. This paper describes the development and implementation of NeuroBlu, the procedures to safeguard data integrity and security and how the data set supports the generation of real-world evidence (RWE) in mental health.

Participants As of 31 July 2021, 562 940 individuals (48.9% men) were present in the data set with a mean age of 33.4 years (SD: 18.4 years). The most frequently recorded diagnoses were substance use disorders (1 52 790 patients), major depressive disorder (1 29 120 patients) and anxiety disorders (1 03 923 patients). The median duration of follow-up was 7 months (IQR: 1.3 to 24.4 months).

Findings to date The data set has supported epidemiological studies demonstrating increased risk of psychiatric hospitalisation and reduced antidepressant treatment effectiveness among people with comorbid substance use disorders. It has also been used to develop data visualisation tools to support clinical decision-making, evaluate comparative effectiveness of medications, derive models to predict treatment response and develop NLP applications to obtain clinical information from unstructured EHR data.

Future plans The NeuroBlu data set will be further analysed to better understand factors related to poor clinical outcome, treatment responsiveness and the development of predictive analytic tools that may be incorporated into the source EHR system to support real-time clinical decision-making in the delivery of mental healthcare services.

Strengths and limitations of this study

- The NeuroBlu data set benefits from a large sample size of deidentified electronic health record (EHR) data from over 560 000 people who have received mental healthcare over a period of 21 years.
- The data set is built on a robust deidentification pipeline and encryption framework that enables a wide range of users to safely analyse data through a trusted research environment using a graphical user interface or advanced analytic software (R and Python).
- Structured data on clinical severity (Clinical Global Impression—Severity (CGI-S)) are recorded for over 80% of patients in the data set. Natural language processing (NLP) enables access to rich clinical data from unstructured free text data as part of the mental state examination and social factors that are not typically available in purely structured clinical data sets such as claims or randomised controlled trial data.
- As the data set draws on real-world EHR data, some variables recording sociodemographic and clinical characteristics are incomplete. However, high completion rates for CGI-S have enabled derived measures to be developed, where certain clinical data are incomplete. When used to extract information from clinical text, NLP models can yield false positives and false negatives. Therefore, downstream analyses of NLP-derived data from EHRs need to take NLP error rates into account.
- Only data from healthcare providers using the MindLinc EHR system are available at present. This means that data on clinical interactions in other acute general medical and primary care settings are not available and recording of comorbid general medical conditions may be incomplete.

INTRODUCTION

Mental disorders contribute to a substantial global burden of illness affecting approximately 1 billion people worldwide.¹ In the USA, around 20% of adults have a mental

disorder with greater prevalence among young adults (29.4%) and women (24.5%).² Around 13 million US adults live with a serious mental illness (SMI) and 65.5% received mental health treatment during 2019.² People with SMI have a markedly reduced life expectancy³ and the lifetime costs associated with SMI are estimated to be \$1.85 million USD per affected individual.⁴ Substance use disorders are associated with considerable disease burden in the USA (1460.3 Disability Adjusted Life Years per 100 000 people).⁵ There are significant barriers to treatment among people with mental disorders. Around 10.3% of US adults with mental illness do not have medical insurance and 57.2% received no treatment in 2020.⁶

The increasing availability of real-world electronic healthcare data sets, known as real-world data (RWD), has improved the understanding of a range of medical disorders at a population level⁷ and contributed to the development⁸ and evaluation⁹ of novel therapeutics. RWD may be analysed to generate real-world evidence (RWE) to quantify the safety and effectiveness of therapeutic interventions. The generation of RWE from the analysis of large-scale data sets that are representative of healthcare service delivery in routine settings provides a complementary source of data—as compared with more controlled clinical trials—which can support the assessment and approval of new treatments.¹⁰ RWE can also be used to guide decision-making for healthcare providers and payers since it more accurately represents how certain treatments are used in routine care and the associated outcomes.¹¹

Electronic health records (EHRs) enable clinicians to document patients' clinical assessment and treatment pathways and represent a rich source of RWD, which could supplement interventional studies and support evidence generation to better characterise health disorders and to develop and deliver more effective treatments.¹² EHR data sets have previously supported RWE generation in critical care¹³ and infrastructure based on common data models¹⁴ has been developed to enable researchers to securely analyse RWD.^{15 16} In principle, such data sources can also be valuable tools to improve access to mental healthcare.

However, unlike physical healthcare, there are considerable challenges in the application of RWE to support mental healthcare knowledge generation.¹⁷ First, mental healthcare data are typically based on clinical evaluation of patients limited to diagnostic codes and subjective, non-standardised clinical impressions. Symptom scales are not recorded in routine clinical practice. Second, there are virtually no reliable and valid objective biomarker data to support the diagnosis or treatment of mental disorders.¹⁸ Third, most clinical data in mental healthcare EHRs are stored as unstructured free text that requires recoding into structured data prior to being analysed. Relatedly, there is little standardisation in EHR data structure between different mental healthcare providers. There is, therefore, a pressing need to improve the availability of

quantifiable RWE derived from mental health EHR data to support a better understanding of factors associated with clinical outcomes, the development and evaluation of more effective treatments and to improve access to timely and effective mental healthcare.

To date, several deidentified mental health EHR data sets have been curated to support large-scale population research,^{19–21} to provide clinical phenotype data for genomics research²² and to develop risk prediction tools with the potential to support clinical decision-making.^{23–26} Natural language processing (NLP) tools have been developed to support automated classification of unstructured free text in EHRs and enable statistical analysis of detailed clinical data such as presenting symptoms,²⁷ environmental factors²⁸ and clinical sentiment.²⁹ NLP infrastructure has previously been applied to unstructured physical health EHR data and made available to researchers through a common data model.³⁰ However, many existing mental health EHR data sets are only accessible by clinical researchers based in specialised academic health science centres.³¹ Enabling wider access, while maintaining robust data privacy and security standards, could substantially increase the impact of deidentified EHR data sets on RWE generation to improve clinical outcomes in people with mental disorders.

This paper presents a cohort profile of the NeuroBlu platform,³² a trusted research environment, which enables safe and secure analysis of deidentified EHR data from US mental healthcare providers, including state-of-the-art NLP software to characterise mental state information from semistructured text,³³ and a graphical user interface to enable users from a wide range of backgrounds to perform statistical analysis to support mental health RWE generation.

COHORT DESCRIPTION

Participants and setting

The NeuroBlu platform enables users to analyse deidentified data from over 560 000 patients receiving care from 25 US mental healthcare providers who operate the MindLinc EHR system. Online supplemental eTable 1 provides details of the locations, types of mental healthcare service and numbers of patients for each provider. In summary, the platform includes deidentified EHR data from individuals receiving mental healthcare from outpatient, inpatient, telemedicine and residential care facilities in 12 US states spanning 21 years (between 1999 and 2020).

The MindLinc EHR system was developed at Duke University Medical Center to enable mental healthcare professionals to document clinical information while providing routine patient care.³⁴ MindLinc includes structured fields to record sociodemographic data, diagnoses, medications and clinical outcome scales as well as semistructured free text fields to document the mental state examination (MSE) and treatment plan.³³ A subset of deidentified MindLinc EHR data are generated (using

the data pipeline described subsequently) to support secondary analyses in NeuroBlu.

Data pipeline

To comply with the Health Insurance Portability and Accountability Act (HIPAA, 1996), MindLinc EHR data are deidentified at source before being transformed and normalised in a cloud-based US Amazon Web Services (AWS) data warehouse and mapped to a common data model to create a harmonised data set, including data from all participating healthcare providers (online supplemental eFigure 1). NeuroBlu users analyse the harmonised data set within a web-based trusted research environment through which analyses may be performed with statistical analysis tools running within a secure cloud-based environment, which does not permit direct access to raw data.³⁵ In this way, deidentified EHR data may be analysed by anyone with access to the trusted research environment without requiring any movement or disclosure of the underlying data.

Deidentification procedure

MindLinc EHR data are deidentified at source (ie, within the computing infrastructure of each participating healthcare provider) using the Safe Harbor method (online supplemental eFigure 1, step 1).³⁶ Where present, this method removes 18 types of identifying information to fully and accurately protect all information that could potentially identify a patient within the data set. This includes names, all geographic subdivisions smaller than a state, all elements of dates that are directly related to an individual (such as birth date, visit date, etc), telephone numbers, vehicle identifiers and serial numbers, fax numbers, device identifiers and serial numbers, email addresses, Uniform Resource Locators (URLs), social security numbers, Internet Protocol (IP) addresses, medical record numbers, biometric identifiers, health plan beneficiary numbers, full-face photographs, account numbers, certificate/license numbers and any other unique identifying number, characteristic or code. These protected items are handled either by outright removal where applicable (such as with names and specific dates) or by transformations to mask the sensitive data while preserving their statistical and analytical value (such as randomising identifiers, recording age in years and collecting the first three digits of zip codes with a population greater than 20 000).

Transfer of deidentified EHR data from healthcare provider to secure AWS cloud

The deidentified EHR data are compressed and transferred from the source healthcare provider through a virtual private network to a secure file transfer protocol site (online supplemental eFigure 1, step 2). The data are then transferred to a secure Amazon Elastic Compute Cloud (AWS EC2) instance (online supplemental eFigure 1, step 3), where they are decompressed and assembled

for loading into a PostgreSQL database (online supplemental eFigure 1, step 4).

Given the relational nature of the MindLinc databases, the files produced by the EHR data deidentification and export process retain a relational structure, storing a patient's data in multiple tables that can be linked together using a unique, randomly generated identifier. The deidentification process is irreversible, preventing reidentification of data, but enabling data from different parts of the database to be joined together at individual patient level. This facilitates the next step of the data pipeline to assemble data from different sources within a Common Data Model (online supplemental eFigure 1, step 4) and is further described in section 1 of the online supplemental material and eFigure 2.

NLP pipeline to extract MSE and social history data

The MSE is a key component of psychiatric assessment during which clinicians assess an individual's clinical presentation at the time of assessment. MSE features may be observed during clinical assessment (eg, appearance and behaviour) or may be elicited through direct questioning or the presentation of cognitive tasks. The MSE provides a rich source of clinical phenotype data which characterises the nature of presenting symptoms, which could be associated with varying clinical outcomes and response to treatment.

The MindLinc EHR includes a semistructured 'status assessment' field in which clinicians can document features associated with a patient's MSE. The status assessment field allows clinicians to choose predefined features from a list of options and/or to document findings as unstructured free text. However, predefined features do not adequately capture the complexity and variability of MSE between different individuals and clinicians largely document MSE features as free text. The median percentage of clinical assessments per patient with documented predefined features in MindLinc is 0%, whereas the median percentage with unstructured free text is 67%.

NLP is the subdiscipline of artificial intelligence that deals with naturally occurring human language, including techniques that enable automated extraction and classification of features from unstructured free text that would be otherwise unfeasible to manually extract by reading through large volumes of text. The application of NLP to mental health EHR data typically involves a series of processes to develop algorithms that can identify clinically meaningful concepts for secondary analysis. These processes include: (1) data assembly: identifying a collection of relevant documents (the corpus), (2) annotation: clinical experts annotate a selection of the corpus to classify meaningful features to generate a training set to develop the NLP model and a reference set to evaluate its performance (3) preprocessing: preparation of the corpus for NLP model development including stop word removal, stemming, lemmatisation and parts of speech tagging, (4) featurisation: classifying text within

the corpus into different features (eg, parts of speech, word vectors or embeddings, sentiment or temporal features) and (5) analysis: development of NLP algorithm using a rule-based or machine learning approach using the training set. The resulting models are evaluated on a previously annotated, held-out reference set and tuned to maximise accuracy as measured through precision (positive predictive value), recall (sensitivity) and F1 measure (harmonic mean of precision and recall).³⁷ If sufficiently accurate models are developed, they can be applied to the entire corpus to generate structured data on clinically meaningful features of interest²⁷ to support mental health RWE generation.^{28 38–42} NLP models have also been investigated as a potential method to screen social media data to identify risk of suicide.⁴³

NLP tools have been previously developed to extract clinical features from unstructured MindLinc MSE data. Details of the NLP pipeline and accuracy statistics have been previously published.³³ In summary, a deep learning, long–short-term memory (LSTM) approach was used to develop NLP applications to extract 241 MSE features in 27 categories. The applications were run over the deidentified EHR data set to create a table of NLP-derived MSE measures, which may be joined and analysed with other structured clinical and sociodemographic data in the NeuroBlu data set. In addition to MSE data, NLP applications have been developed to extract data on environmental stressors as part of the social history. Further information on the structure of NLP-derived data is provided in the Unstructured data section.

Data variables

Full details of available data variables and their structure within the NeuroBlu MindLinc data set are provided in section 2 of the online supplemental material. The data variable structure is derived from the structure of the source MindLinc EHR and transformed into a relational database, which may be accessed using predefined database queries in the front-end interface or through embedded SQL queries in R scripts (described subsequently).

Structured data

The data set includes structured sociodemographic data on gender, year of birth, race, ethnicity, US state of residence, marital status, employment status and educational history. Sociodemographic data may be joined to structured clinical data represented in other tables within the data set that include data on contacts with mental health services, emergency room visits, hospital admissions, International Classification of Diseases ICD-9/ICD-10 diagnosis, prescribed medications and recorded family history of mental disorders. Data from structured rating scales are also available, including Clinical Global Impression—Severity (CGI-S),⁴⁴ Clinical Global Impression—Improvement (CGI-I),⁴⁴ Global Assessment of Functioning (GAF)⁴⁵ and Montgomery-Asberg Depression Rating Scale (MADRS).⁴⁶

As structured rating scales are applied at clinicians' discretion, they are not completed at every patient encounter and this limits their availability within the NeuroBlu MindLinc data set. Structured questionnaire data are available for the following numbers of patients within the data set:

1. CGI-S: 471 256 patients (83.7%).
2. CGI-I: 361 819 patients (64.3%).
3. GAF: 310 895 patients (55.3%).
4. MADRS: 1126 patients (0.2%).

The data set benefits from high completion rates of CGI-S but lower completion rates of other rating scales. Furthermore, CGI-S data are not always recorded at every patient encounter and the score may fluctuate, making it difficult to interpret. To address these limitations, derived scores have been implemented to estimate the total MADRS score (based on correlation with CGI-S) and to smooth CGI-S data through interpolation. This technique enables the MADRS score to be estimated when it is not recorded and the CGI-S score to be estimated during clinical encounters where it is not recorded and where multiple scores are recorded during the same visit. Details of these methods are provided in section 2 (iii) of the online supplemental material.

Unstructured data

In addition to the structured sociodemographic and clinical described above, NLP applications (described previously) have been applied to extract data on social history and MSE from unstructured free text recorded by clinicians in the MindLinc EHR. Full details on the NeuroBlu NLP applications are provided in section 2(i) (external stressors/social history) and 2(iv) (MSE) of the online supplemental material. Information on NLP model development and error rates for MSE categories are also provided in the user guide on the NeuroBlu platform.

In summary, data available on external stressors as part of social history include difficulties with social and family relationships, major life events, alleged history of abuse, environmental problems, financial difficulties, problems with housing, forensic history, problems at school and difficulties in employment. These factors have been shown to be associated with the risk of developing a mental disorder and provide important context to the relationship between other sociodemographic and clinical factors with outcomes.⁴⁷ To maintain data security and confidentiality, the original unstructured EHR data are not available for analysis. For social history/external stressors and MSE data, only the data derived using NLP are available for analysis in the NeuroBlu platform.

NLP-derived MSE data are represented as 241 binary variables (indicating the presence or absence of a particular feature) as part of 27 categories: abnormal or psychotic thoughts; affect; appearance; association; attention/concentration; attitude; cognition; executive functioning; fund of knowledge; gait and station; homicidal; impulse control; insight; intelligence; judgement; language; level of consciousness; memory; mood;

orientation; psychomotor; reasoning; sensorium; sleep; speech; suicidal and violent thoughts. The default category for each value is, 'no issues' (or equivalent) and represents the most frequent label. Where a clinician documents an abnormality, this is represented by a specific descriptor within a particular category which is derived from unstructured text using NLP. More than one variable may be present within each category to describe multiple clinical features.

NLP-derived data may be joined at an individual patient level using the date that the data were documented. This enables analysis of unstructured data and structured data recorded on the same date or within a particular time window (eg, within 14 days of the recorded diagnosis date or following admission to hospital). Using this approach, it is possible to generate a rich clinical phenotype of social factors and presenting mental state, which can be analysed against treatment exposure and clinical outcomes. For example, selecting a cohort of patients with a diagnosis of schizophrenia and prominent cognitive or negative symptoms who may have a different phenotype and response to treatment compared with patients with prominent positive psychotic symptoms.⁴⁸ The use of such data could help to develop a better understanding of underlying pathophysiology beyond traditional diagnostic classification and identify patient groups who may benefit from emerging treatment that targets specific symptom domains.

NeuroBlu front-end interface

Authenticated users can access NeuroBlu through a secure web-based interface (<https://app.neuroblu.ai/>). The interface contains four key elements: (1) Cohort Builder, (2) Category Mapper, (3) Data Explorer and (4) R/Python Code Engine. These elements allow users to define a population cohort based on specified inclusion/

exclusion criteria, define key variables to examine within the cohort using descriptive statistics and charts and perform more detailed inferential statistics and predictive analytics using R or Python, open-source statistical analysis software packages that are frequently used in healthcare data research.

Cohort Builder

In the Cohort Builder, users can assign inclusion and exclusion criteria to filter the data set into a specific cohort (figure 1). The criteria may include age at first clinical contact, gender, race, exposure to psychiatric medications and psychiatric diagnosis. As inclusion/exclusion criteria are applied, descriptive statistics on sociodemographics and availability of data over time and clinical severity (CGI-S and proportion hospitalised) are dynamically updated. This enables users to rapidly assess the population characteristics and availability of data for a particular cohort.

Category Mapper

The Category Mapper allows users to recode individual variables into new categories, which can subsequently be called in statistical analysis using R/Python Code. For example, continuous age may be recoded as discrete age bands and individual medications can be recoded into larger categories (online supplemental eFigure 3). This enables users to predefine variables for statistical analysis without having to repeat the same code to compute new variables in every R/Python script. The same mappings can be used for separate cohorts to ensure reproducibility of statistical methods in different analyses.

Data Explorer

A summary of the entire data set is provided in the Data Explorer with accompanying Dataset Specifications that

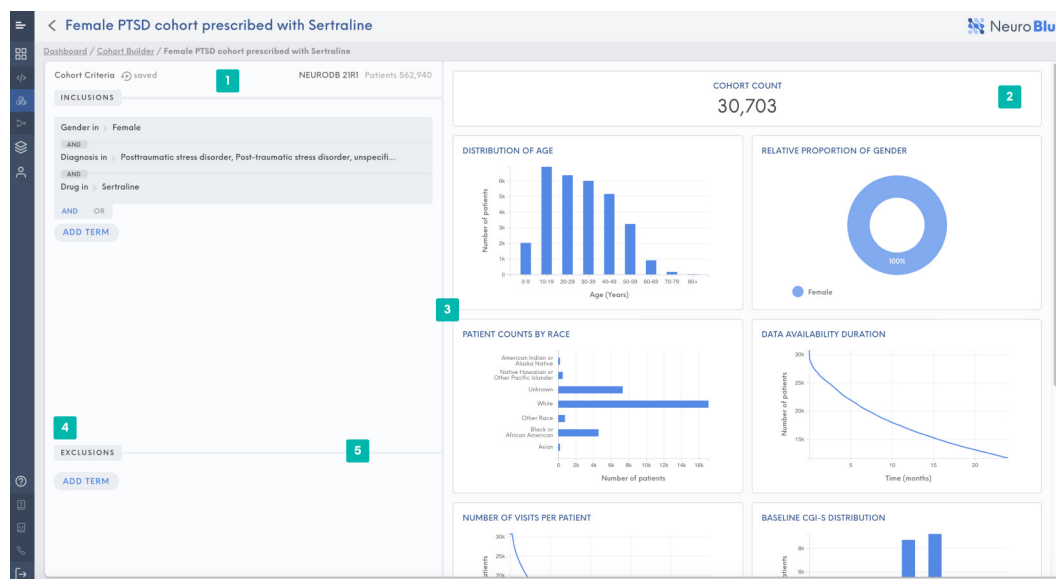


Figure 1 NeuroBlu Cohort Builder illustrating the population characteristics of female patients diagnosed with post-traumatic stress disorder (PTSD) who have received sertraline.

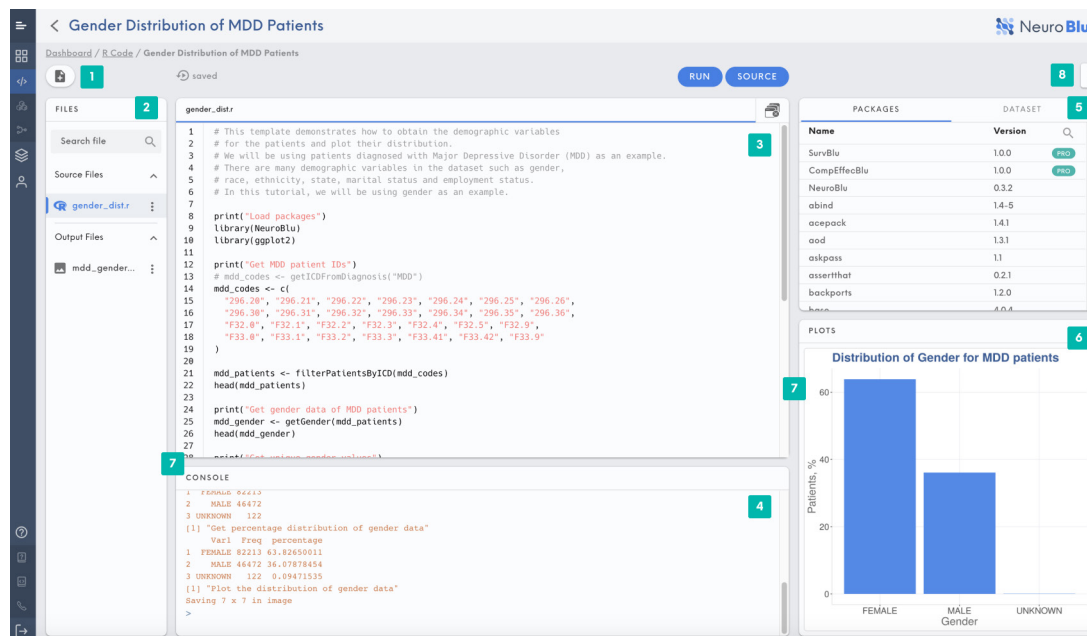


Figure 2 NeuroBlu R Code Engine includes a code editor, console, file manager and output viewer to perform data assembly and statistical analyses. An analogous Python Code Engine is also available. MDD: Major Depressive Disorder

describe the data structure (online supplemental eFigure 4). The top 50 rows of each table can be viewed within the Data Explorer to allow users to understand the data structure and elements within each row to facilitate development of table joins to assemble clinical data at individual patient level. However, it is not possible to export or download any part of the data set from the secure web-based interface.

R/Python Code Engine

NeuroBlu enables users to perform custom data assembly and analytics with an R Code/Python Engine (figure 2). The web-based platform includes a *code editor* to build a data assembly/analysis pipeline, a *console* to view the execution of R/Python code, a *file manager* to manage and execute different scripts within the analysis project, and an *output viewer* which provides the tabular and graphical outputs of R/Python code analysis. Previously defined cohorts and variables from the Cohort Builder and Category Mapper can be directly called into analysis scripts. Outputs of statistical analyses can be exported as comma-separated values (CSV) files (tables) or PNG/JPG files (graphs). The R/Python Code Engine includes commonly used R/Python packages as well as three custom R/Python packages designed to support analysis of NeuroBlu data: (1) *NeuroBlu*: utility functions which simplify data assembly processes using prebuilt SQL queries, (2) *SurvBlu*: a framework for conducting time-to-event analyses and (3) *CompEffecBlu*: a collection of functions designed to conduct comparative effectiveness research studies on medication data. NeuroBlu also provides several template scripts to facilitate commonly performed data procedures such as assembling and visualising data.

Data privacy

Under HIPAA regulations, US healthcare providers record and store personal healthcare information for the purposes of providing treatment, billing data, to support healthcare service operations. Deidentified data may be used for secondary analyses.

Data security

The NeuroBlu platform's data security framework complies with Center for Internet Security (CIS) benchmarks, which are documented industry best practices for securely configuring IT systems, software and networks. The platform complies with a CIS level 2 for Operating System (OS) level security in deployment, which is managed and enforced through InSpec. For data encryption, all data in-transit through the platform is encrypted with TLS1.2 and all data at rest are encrypted with AES256.

To ensure that the NeuroBlu account and its access are secure, the NeuroBlu platform was developed following a secure software development life cycle process. The process is defined and enforced via an ISO27001-certified information security management system. User access authentication is maintained through an access-control list enforced through the platform application.

Participant characteristics

Descriptive statistics for the NeuroBlu data set were obtained on 31 July 2021. A total of 562 940 individuals (48.9% male) were present in the data set with a mean age of 33.4 years (SD: 18.4 years). The majority of the population was white (49.1%). Online supplemental eTable 2 and eFigure 5 provide a full breakdown of the population by race. Children and adolescents up to 19 years of age account for 28.7% of the data set, working

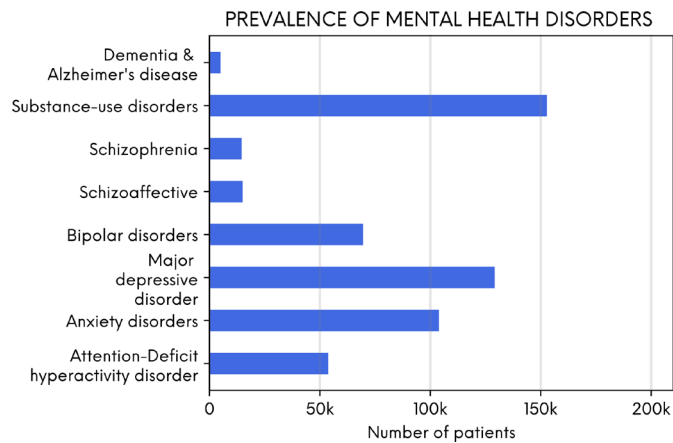


Figure 3 Prevalence of mental disorders in the NeuroBlu data set.

age adults (aged 20–59 years) account for 62.2% and adults aged 60 years or above account for 9.1% (online supplemental eTable 3/eFigure 6).

Online supplemental eTable 4 and figure 3 illustrate the prevalence of mental disorders in the data set. The most frequently recorded diagnoses were substance use disorders (28.1%) followed by major depressive disorder (23.7%) and anxiety disorders (19.1%). The gender distribution of diagnoses is provided in online supplemental eTable 5 and eFigure 7. These data show a greater frequency of substance-use disorders, schizophrenia and attention-deficit hyperactivity disorder (ADHD) among male patients and a greater frequency of dementia and Alzheimer's disease, bipolar disorder, major depressive disorder and anxiety disorders among female patients. The constituent diagnostic codes are provided in online supplemental eTable 6.

Online supplemental eTable 7 and eFigure 8 provide a breakdown by US State of Residence. The most frequent US States of Residence were Colorado (23.0%), New York (18.9%), Missouri (18.1%) and North Carolina (13.9%).

The mean CGI-S score recorded across all clinical visits by diagnosis is provided in online supplemental eTable 8 and eFigure 9. The greatest mean CGI-S scores were found in people with schizoaffective disorder (4.48, SD: 1.42) and schizophrenia (4.44, SD: 1.44) with the lowest mean score recorded in people with substance-use disorders (3.61, SD: 1.80). A further analysis was conducted among patients with at least two recorded CGI-S scores to estimate the mean maximum and mean minimum score by diagnosis (eTableonline supplemental eTable 9 and figure 4). The difference between the mean maximum and minimum CGI-S score was lowest in people with dementia and Alzheimer's disease. This may reflect the pervasive and progressive nature of dementia compared with mood, anxiety and psychotic disorders, which are typically characterised by periods of remission in between episodes.

The number of patients with at least one hospital visit (inpatient or emergency department) by diagnosis (eTableonline supplemental eTable 10 and eFigure 10)

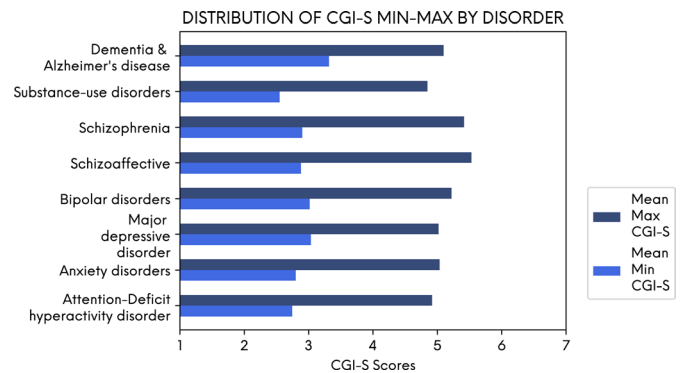


Figure 4 Mean maximum and mean minimum Clinical Global Impression—Severity (CGI-S) score by diagnosis.

was greatest for patients with schizoaffective disorder (56.7%), schizophrenia (55.7%) and dementia and Alzheimer's disease (52.2%) reflecting the severity of these disorders leading to increased risk of psychiatric hospitalisation compared with other mental disorders.

NLP-derived MSE data were obtained in three clinical domains: delusions & hallucinations, mood and cognition. These domains were chosen as they represent key clinical features, which may be experienced in a wide range of mental disorders. The constituent NLP applications for each domain are defined in online supplemental eTable 11. Three variables were created for each domain defined as the presence of at least one feature within each domain at any point during an individual's clinical record. The breakdown of these variables by diagnosis is provided in online supplemental eTable 12 and figure 5. Delusions and hallucinations were most frequently recorded in people with schizophrenia, schizoaffective disorder, bipolar disorder and dementia and Alzheimer's disease, in keeping with the impact of these disorders on thought content and perception. Mood features were widely documented across all mental disorders but to a lesser extent among people with dementia and Alzheimer's disease. Features related to cognition were most frequently documented in people with dementia and Alzheimer's disease as well as people with schizophrenia and schizoaffective disorder, reflecting the importance of cognitive symptoms in these disorders.

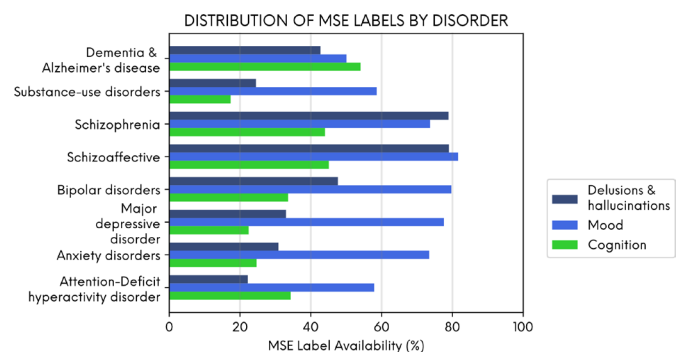


Figure 5 Percentage of patients with at least one natural language processing (NLP)-derived mental state examination (MSE) feature by diagnosis.

Online supplemental eFigure 11 provides a distribution of the total number of clinical encounters (visits) per patient. Patients had a median of five visits recorded (IQR: 2–20 visits). Online supplemental eFigure 12 provides a distribution of the total follow-up duration available for each patient in the data set. The median follow-up duration was 7 months (IQR: 1.3–24.4 months).

Data were available for age in all patients. Data were missing for the following variables:

1. Gender: 1572 patients (0.23%).
2. Race: 1 865 17 patients (33.1%).
3. Psychiatric diagnosis: 66 064 patients (11.7%).
4. State: 21 919 patients (3.9%).
5. NLP-derived MSE data: 1 326 90 patients (23.6%).

Patient and public involvement

Patients were not involved in the development of the data set or reporting of the cohort profile.

Findings to date

The NeuroBlu platform and MindLinc EHR data set have been used to generate RWE through large-scale mental health epidemiology studies as well as to develop automated NLP, data visualisation and predictive analytic tools based on real-world EHR data.

Epidemiology of mental disorders

Data from people with substance use disorders has been analysed to investigate its impact among minority ethnic people⁴⁹ and demonstrated increased rates of psychiatric hospitalisation among children and adolescents⁵⁰ and increased risk of psychiatric comorbidities and hospitalisation among adults with substance use disorders.⁵¹

The availability of structured clinical rating scale data has enabled analysis of the associations of psychotropic medications with clinical outcomes. An analysis of antidepressant monotherapy in people with depression found a significant reduction in CGI-S following treatment but less improvement seen in people with comorbid substance use or anxiety disorders.⁵² Augmentation of antidepressant treatment with a second-generation antipsychotic in people with depression was more likely in those with a high baseline CGI-S.⁵³

Longitudinal prescribing data have been analysed to demonstrate that antidepressant prescribing reduced in response to US Food and Drug Administration (FDA) warnings on risk of suicidality in children and adolescents.⁵⁴ The duration of continued treatment with buprenorphine (an opioid receptor partial agonist used to treat people with opioid use disorder) is shorter in people with greater opioid use disorder severity and in people with comorbid cannabis use disorder.⁵⁵

Mental health informatics

As well as being analysed to generate RWE, the data set has been used to develop novel data visualisation and predictive analytic methods to support the analysis of mental health RWD including a data visualisation tool to support clinical decision making,⁵⁶ a model to predict

treatment response in major depressive disorder,⁵⁷ infrastructure to augment data from clinical trials with clinician-administered rating scales from the EHR,⁵⁸ and a data visualisation tool to evaluate comparative effectiveness of psychotropic medications.⁵⁹

Furthermore, the availability of unstructured free text data has supported the development of NLP tools to automatically extract clinically meaningful information from the data set, as described previously.^{33 60}

Strengths and limitations

The NeuroBlu data set contains a rich source of deidentified EHR data that supports the generation of RWE to improve understanding of mental disorders and factors affecting clinical outcomes. NeuroBlu includes data on over 560 000 people attending 25 US mental healthcare providers over a period of 21 years. The large sample size and long duration of follow-up enable well-powered statistical analyses that are representative of real-world clinical practice. This is a key strength that allows the investigation of research questions that are not feasible to address with randomised controlled trials or prospective observational studies (eg, comparative effectiveness studies on multiple medications or assessing the impact of substance use disorders on clinical outcomes).

NeuroBlu is built on a robust deidentification pipeline that ensures no personal identifiable data are available to users who analyse the data. A strong data security framework that employs internationally validated encryption standards protects the anonymised data set from being accessed outside the trusted research environment of NeuroBlu. These safeguards ensure the security of the data set while still enabling users to easily analyse the data through a secure web-based interface. The graphical user interface of NeuroBlu contains built-in features to define population cohorts for analysis (Cohort Builder), recode data variables (Category Mapper) and visualise the distribution of clinical data (Data Explorer). In addition, embedded statistical software (including R and Python Code Engines) enable users to directly query, assemble and analyse the data while ensuring the data remain secure within the trusted research environment. These features allow users from a wide range of backgrounds to analyse the data without requiring expertise in statistical software while at the same time enabling expert data scientists to perform advanced analytical procedures on the data set using R or Python.

The NeuroBlu infrastructure has been developed using the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM), a standardised approach to harmonise healthcare data sets. This approach will be used to introduce new anonymised mental healthcare data sets into the NeuroBlu trusted research environment using a federated data query engine to further increase the availability of data within the platform.

The data set benefits from structured sociodemographic data as well as clinician-rated symptom scales

(CGI-S, CGI-S, GAF and MADRS). CGI-S data are available for over 80% of patients in the data set allowing statistical analyses to be performed against a validated marker of clinical severity. Another strength is the use of NLP to obtain clinical data from unstructured free text on clinical features as part of the MSE as well as data on social factors. Together with the available structured data, these provide a detailed clinical phenotype of individuals receiving mental healthcare that may support the analysis of associations of treatments with different clinical presentations. This approach is typically unfeasible when analysing EHR data sets containing only structured data or claims data. However, a limitation of NLP models is that they can yield false-positive and false negative instances that can introduce errors into secondary analyses of NLP-derived data from EHRs. NLP-derived data also depend on the presence of documented clinical information in free text records. The absence of documentation does not necessarily indicate the absence of a particular clinical construct and clinicians do not systematically document the absence of clinical features unless it is clinically relevant to do so.

As EHR data are recorded as part of routine clinical practice, a limitation of the data set is that clinical data may not be comprehensively completed across all domains for all patients. Clinicians typically record data that are relevant to an individual's clinical presentation or treatment plan but will not necessarily document the absence of non-relevant clinical features and may not always complete structured clinical rating scales. Notably, the completion rate for the MADRS scale is less than 1%. However, this limitation has been addressed by computing a derived score for MADRS based on available CGI-S data. Furthermore, the availability of rich clinical data within unstructured text (extracted using NLP) combined with clinical severity data from the CGI-S means that it is possible to develop derived measures on symptom severity where structured symptom rating scales are unavailable.

At present, the data set only includes clinician-recorded EHR data and does not include any patient-reported outcome measures (PROMs) such as patient-recorded symptom or medication side effect scales. PROMs collected in between periods of clinical contact could help to address the limitations of EHR data described previously by providing patient-rated data on mental health symptom burden and response to treatment that could support more nuanced evidence generation and clinical decision-making.

The data set includes data on prescribed medications, start dates, stop dates and dosage information. However, it is not possible to determine adherence to treatment as this is not typically directly observed in routine clinical practice. Additionally, where an individual is receiving care from more than one healthcare provider, data on prescribed medications from other providers may not be available.

As the data set reflects a dynamic population, patients may have received care from other healthcare providers

or move into or out of the catchment population of clinics using the MindLinc EHR system. This means that data relating to healthcare outside of MindLinc healthcare providers (eg, acute physical healthcare providers or diagnoses and medications recorded in community primary care settings) are not available within the NeuroBlu data set.

FUTURE PLANS

As patients continue to receive care from healthcare providers using the MindLinc EHR, the NeuroBlu data set will grow over time increasing in sample size and follow-up duration. We plan to analyse these data to generate RWE to better understand the factors related to poor mental health outcomes, to evaluate the impact of treatments, to develop more advanced NLP models with improved classification accuracy in collaboration with experts in this area and to develop predictive analytic tools that quantify these factors at individual patient level.⁶¹ We aim to incorporate these tools into the source EHR alongside patient-reported outcome data to provide clinicians with actionable insights to support real-time clinical decision-making. This approach could help to improve clinical outcomes by better personalising mental healthcare and reducing delays to effective treatment.

COLLABORATION

The robust data security framework of the NeuroBlu platform and the availability of embedded data query and statistics tools within a web-based graphical user interface enable the MindLinc data set to be analysed by a wide range of users from varied professional disciplines in any geographical location. We actively collaborate with academic partners and mental healthcare providers to support mental health RWE generation.⁶² To this end, we highly encourage collaboration with individuals, academic institutes, healthcare providers and commercial organisations who wish to analyse the data set to support mental health research and to inform healthcare policy and clinical practice. We would also welcome collaboration with healthcare providers who would be interested in using the NeuroBlu deidentification pipeline and trusted research environment to support RWE generation from their own data sets. Further information on the NeuroBlu platform is available on our website (<https://www.neuroblu.ai>) and please do contact us at info@neuroblu.ai to discuss opportunities for collaboration and access to the platform.

Twitter Rashmi Patel @RPatelDr

Contributors The manuscript was conceived by RP and SK. Data assembly and statistical analysis were performed by SNW and RR. Reporting of findings was carried out by RP with support from SK. All authors (RP, SNW, RR, ST, JT, RG, NC, MV, AJR, MER, JS and SK) contributed to cohort profile design, manuscript preparation and approved the final version. SK is guarantor.

Funding The NeuroBlu data set and MindLinc EHR system are owned and funded in full by Holmusk.

Competing interests RP has received funding from the National Institute of Health Research (NIHR301690), the Medical Research Council (MR/S003118/1), the Academy of Medical Sciences (SGL015/1020) and Janssen. AJR has received consulting fees from Compass Inc., Curbstone Consultant LLC, Emmes Corp., Evecxia Therapeutics, Inc., Holmusk, Liva-Nova, Neurocrine Biosciences Inc., Otsuka-US, speaking fees from Liva-Nova, Johnson and Johnson (Janssen) and royalties from Guildford Press. The other authors declare no competing interests.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants but The NeuroBlu platform has received a waiver of authorisation for analysis of deidentified healthcare data from the WCG Institutional Review Board (Ref: WCG-IRB 1-1470336-1).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. Deidentified data may be analysed through the NeuroBlu platform (<https://www.neuroblu.ai/>). We highly encourage collaboration with individuals, academic institutes, healthcare providers and commercial organisations who wish to analyse the data set to support mental health research and to inform healthcare policy and clinical practice. Please e-mail info@neuroblu.ai for further information on accessing NeuroBlu.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Rashmi Patel <http://orcid.org/0000-0002-9259-8788>

REFERENCES

- Rehm J, Shield KD. Global burden of disease and the impact of mental and addictive disorders. *Curr Psychiatry Rep* 2019;21:10.
- Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: results from the 2019 national survey on drug use and health, 2020. Available: <https://www.samhsa.gov/data/sites/default/files/reports/rpt29393/2019NSDUHFFRPDFWHTML/2019NSDUHFFR1PDF/W090120.pdf>
- Colton CW, Manderscheid RW. Congruencies in increased mortality rates, years of potential life lost, and causes of death among public mental health clients in eight states. *Prev Chronic Dis* 2006;3:1-14.
- Seabury SA, Axteen S, Pauley G, et al. Measuring the lifetime costs of serious mental illness and the mitigating effects of educational attainment. *Health Aff* 2019;38:652-9.
- GBD 2016 Alcohol and Drug Use Collaborators. The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990-2016: a systematic analysis for the global burden of disease study 2016. *Lancet Psychiatry* 2018;5:987-1012.
- Mental health America. 2020 access to care data. MENT. heal. am, 2020. Available: <https://www.mhanational.org/issues/2020/mental-health-america-access-care-data#four>
- Maissenhaelter BE, Woolmore AL, Schlag PM. Real-World evidence research based on big data: Motivation-challenges-success factors. *Onkologie* 2018;24:91-8.
- Singh G, Schulthess D, Hughes N, et al. Real world big data for clinical research and drug development. *Drug Discov Today* 2018;23:652-60.
- Corrigan-Curay J, Sacks L, Woodcock J. Real-World evidence and real-world data for evaluating drug safety and effectiveness. *JAMA* 2018;320:867-8.
- Klonoff DC. The new FDA real-world evidence program to support development of drugs and biologics. *J Diabetes Sci Technol* 2020;14:345-9.
- Roberts MH, Ferguson GT. Real-World evidence: bridging gaps in evidence to guide payer decisions. *Pharmacoecon Open* 2021;5:3-11.
- Bartlett VL, Dhruva SS, Shah ND, et al. Feasibility of using real-world data to replicate clinical trial evidence. *JAMA Netw Open* 2019;2:e1912869.
- Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016;3:160035.
- OMOP common data model – OHDSI, 2022. Available: <https://www.ohdsi.org/data-standardization/the-common-data-model/>
- NORC.org. NORC data Enclave, 2022. Available: <https://www.norc.org/Research/Capabilities/Pages/data-enclave.aspx>
- Administrative data research facility (ADRF) – Coleridge initiative, 2022. Available: <https://coleridgeinitiative.org/adrf/>
- Stewart R, Davis K. 'Big data' in mental health research: current status and emerging possibilities. *Soc Psychiatry Psychiatr Epidemiol* 2016;51:1055-72.
- Peterson BS. Editorial: biomarkers in precision medicine for mental illnesses. *J Child Psychol Psychiatry* 2020;61:1279-81.
- Perera G, Broadbent M, Callard F, et al. Cohort profile of the South London and Maudsley NHS Foundation trust biomedical research centre (SLAM BRC) case register: current status and recent enhancement of an electronic mental health Record-derived data resource. *BMJ Open* 2016;6:e008721.
- Werbeloff N, Osborn DPJ, Patel R, et al. The Camden & Islington Research Database: Using electronic mental health records for research. *PLoS One* 2018;13:e0190703.
- Cardinal RN. Clinical records anonymisation and text extraction (CRATE): an open-source software system. *BMC Med Inform Decis Mak* 2017;17:50.
- Smoller JW. The use of electronic health records for psychiatric phenotyping and genomics. *Am J Med Genet B Neuropsychiatr Genet* 2018;177:601-12.
- Holderness E, Miller N, Cawkwell P, et al. Analysis of risk factor domains in psychosis patient health records. *J Biomed Semantics* 2019;10:19.
- Bhattacharyya S, Schoeler T, Patel R, et al. Individualized prediction of 2-year risk of relapse as indexed by psychiatric hospitalization following psychosis onset: model development in two first episode samples. *Schizophr Res* 2021;228:483-492.
- Oliver D, Spada G, Colling C, et al. Real-World implementation of precision psychiatry: Transdiagnostic risk calculator for the automatic detection of individuals at-risk of psychosis. *Schizophr Res* 2021;227:52-60.
- Raket LL, Jaskolowski J, Kinon BJ, et al. Dynamic electronic hEalth reCord deTecton (detect) of individuals at risk of a first episode of psychosis: a case-control development and validation study. *Lancet Digit Health* 2020;2:e229-39.
- Jackson RG, Patel R, Jayatilake N, et al. Natural language processing to extract symptoms of severe mental illness from clinical text: the clinical record interactive search comprehensive data extraction (CRIS-CODE) project. *BMJ Open* 2017;7:e012012.
- Irving J, Colling C, Shetty H, et al. Gender differences in clinical presentation and illicit substance use during first episode psychosis: a natural language processing, electronic case register study. *BMJ Open* 2021;11:e042949.
- Holderness E, Cawkwell P, Bolton K. Distinguishing Clinical Sentiment: The Importance of Domain Adaptation in Psychiatric Patient Health Records. In: *Proceedings of the 2nd clinical natural language processing workshop*. Minneapolis, Minnesota, USA: Association for Computational Linguistics, 2019: 117-23.
- Liu S, Wang Y, Wen A, et al. Implementation of a cohort retrieval system for clinical data Repositories using the observational medical outcomes partnership common data model: proof-of-concept system validation. *JMIR Med Inform* 2020;8:e17376.
- Ford T, Mansfield KL, Markham S, et al. The challenges and opportunities of mental health data sharing in the UK. *Lancet Digit Health* 2021;3:e333-6.
- Holmusk. NeuroBlu: the evidence footprint. building the future of behavioral health with real-world data and analytics, 2021. Available: <https://www.neuroblu.ai/>
- Mukherjee SS, Yu J, Won Y, et al. Natural language Processing-Based Quantification of the mental state of psychiatric patients. *Comput Psychiatr* 2020;4:76.
- Gersing K, Krishnan R. Clinical computing: clinical management research information system (CRIS). *Psychiatr Serv* 2003;54:1199-200.

- 35 HDR UK. Response to the Goldacre Review - "Trusted Research Environments and Data Management", 2021. Available: <https://www.hdruk.ac.uk/news/response-to-the-goldacre-review-trusted-research-environments-and-data-management/>
- 36 Kayaalp M. Modes of de-identification. *AMIA Annu Symp Proc* 2017;2017:1044–50.
- 37 Henry S, Yetisgen M, Uzuner O. Chapter 13: Natural Language Processing in Mental Health Research and Practice. In: Tenenbaum JD, Ranallo PA, eds. *Mental health informatics: enabling a learning mental healthcare system*. Cham: Springer International Publishing, 2021: 317–53.
- 38 Patel R, Jayatilake N, Broadbent M, et al. Negative symptoms in schizophrenia: a study in a large clinical sample of patients using a novel automated method. *BMJ Open* 2015;5:e007619.
- 39 Downs J, Dean H, Lechler S, et al. Negative symptoms in early-onset psychosis and their association with antipsychotic treatment failure. *Schizophr Bull* 2019;45:69–79.
- 40 Patel R, Lloyd T, Jackson R, et al. Mood instability is a common feature of mental health disorders and is associated with poor clinical outcomes. *BMJ Open* 2015;5:e007504.
- 41 McDonald K, Smith T, Broadbent M, et al. Prevalence and incidence of clinical outcomes in patients presenting to secondary mental health care with mood instability and sleep disturbance. *Eur Psychiatry* 2020;63:e59.
- 42 Irving J, Patel R, Oliver D, et al. Using natural language processing on electronic health records to enhance detection and prediction of psychosis risk. *Schizophr Bull* 2021;47:405–14.
- 43 Coppersmith G, Leary R, Crutchley P, et al. Natural language processing of social media as screening for suicide risk. *Biomed Inform Insights* 2018;10:117822261879286.
- 44 Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. *Psychiatry* 2007;4:28–37.
- 45 Diagnostic and statistical manual of mental disorders : DSM-IV. Fourth edition. Washington, DC : American Psychiatric Association, 1994. Available: <https://search.library.wisc.edu/catalog/999733358502121>
- 46 Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;134:382–9.
- 47 Allen J, Balfour R, Bell R, et al. Social determinants of mental health. *Int Rev Psychiatry* 2014;26:392–407.
- 48 Carbon M, Correll CU. Thinking and acting beyond the positive: the role of the cognitive and negative symptoms in schizophrenia. *CNS Spectr* 2014;19 Suppl 1:38–52.
- 49 Wu L-T, Blazer DG, Gersing KR, et al. Comorbid substance use disorders with other axis I and II mental disorders among treatment-seeking Asian Americans, native Hawaiians/Pacific Islanders, and mixed-race people. *J Psychiatr Res* 2013;47:1940–8.
- 50 Wu L-T, Gersing K, Burchett B, et al. Substance use disorders and comorbid axis I and II psychiatric disorders among young psychiatric patients: findings from a large electronic health records database. *J Psychiatr Res* 2011;45:1453–62.
- 51 Wu L-T, Gersing KR, Swartz MS, et al. Using electronic health records data to assess comorbidities of substance use and psychiatric diagnoses and treatment settings among adults. *J Psychiatr Res* 2013;47:555–63.
- 52 Lee T-S, Assam PN, Gersing KR, et al. The effectiveness of antidepressant monotherapy in a naturalistic outpatient setting. *Prim Care Companion CNS Disord* 2012;14.
- 53 Gersing KR, Sheehan JJ, Burchett B, et al. Use of augmentation agents for treating depression: analysis of a psychiatric electronic medical record data set. *Psychiatr Serv* 2014;65:1062–5.
- 54 Gupta S, Gersing KR, Erkanli A, et al. Antidepressant regulatory warnings, prescription patterns, suicidality and other aggressive behaviors in major depressive disorder and anxiety disorders. *Psychiatr Q* 2016;87:329–42.
- 55 Balani A, Ker S, Hsu J, et al. T4. buprenorphine treatment retention in individuals with opioid use disorder: insights from real-world dataset. in: American Society of clinical psychopharmacology, 2021. Available: <https://pmg.joynadmin.org/documents/1005/60ad2955ef0d26cf5e684539.pdf#page=181>
- 56 Mane KK, Owen P, Schmitt C. Visual analytics to optimize patient-population evidence delivery for personalized care. In: *2013 ACM conference on bioinformatics, computational biology and biomedical informatics, ACM-BCB 2013. Association for computing machinery*, 2013: 484–8.
- 57 Akushevich I, Kravchenko J, Gersing K. Predictive model of the treatment effect for patients with major depressive disorder. In: *2013 ACM conference on bioinformatics, computational biology and biomedical informatics, ACM-BCB 2013. Association for computing machinery*, 2013: 518–24.
- 58 Pan J-J, Nahm M, Wakim P, et al. A centralized informatics infrastructure for the National Institute on drug abuse clinical trials network. *Clin Trials* 2009;6:67–75.
- 59 Mane KK, Bizon C, Schmitt C, et al. VisualDecisionLinc: a visual analytics approach for comparative effectiveness-based clinical decision support in psychiatry. *J Biomed Inform* 2012;45:101–6.
- 60 Mukherjee S, Won Y, McClay M, et al. W35. natural language processing-based quantification of the mental state of psychiatric patients. in: American Society of clinical psychopharmacology, 2021. Available: <https://pmg.joynadmin.org/documents/1005/60ad2955ef0d26cf5e684539.pdf#page=151>
- 61 Hahn T, Nierenberg AA, Whitfield-Gabrieli S. Predictive analytics in mental health: applications, guidelines, challenges and perspectives. *Mol Psychiatry* 2017;22:37–43.
- 62 Holmusk. NeuroBlu Award 2021, 2021. Available: <https://www.holmusk.com/news/neuroblu-award>