BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare analytics with real-world data

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-057227
Article Type:	Cohort profile
Date Submitted by the Author:	09-Sep-2021
Complete List of Authors:	Patel, Rashmi; Holmusk USA Inc; King's College London Institute of Psychiatry Psychology and Neuroscience, Department of Psychosis Studies Wee, Soon Nan; Holmusk USA Inc Ramaswamy, Rajagopalan; Holmusk USA Inc Thadani, Simran; Holmusk USA Inc Tandi, Jesisca; Holmusk USA Inc Garg, Ruchir; Holmusk USA Inc Calvanese, Nathan; Holmusk USA Inc Valko, Matthew; Holmusk USA Inc Rush, A.; Curbstone Consultant LLC Rentería, Miguel; Holmusk USA Inc Sarkar, Joydeep; Holmusk USA Inc Kollins, Scott; Holmusk USA Inc; Duke University School of Medicine
Keywords:	PSYCHIATRY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, EPIDEMIOLOGY

SCHOLARONE™
Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

For peer review only

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment 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] and Scott H. Kollins[1,4]

1. Holmusk, 54 Thompson Street, New York, NY 10012, USA
2. King's College London (Institute of Psychiatry, Psychology and Neuroscience), London, UK
3. Curbstone Consultant LLC, Santa Fe, NM, USA
4. Duke University School of Medicine (Department of Psychiatry and Behavioral Sciences), Durham, NC, USA

Correspondence to:

Rashmi Patel

E-mail: rashmi.patel@holmusk.com

Manuscript word count (excluding references): 5,137

Abstract word count: 298

For peer review only

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare real-world data analytics

Abstract

Purpose:

NeuroBlu is a real-world data (RWD) repository that contains de-identified electronic health record (EHR) data from U.S. 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, ICD-9/ICD-10 diagnosis, prescribed medications, family history of mental disorders, Clinical Global Impression (CGI-S/CGI-I), and Global Assessment of Functioning (GAF). To further enhance the dataset, 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 dataset supports the generation of real-world evidence (RWE) in mental health.

Participants:

As of 31st July 2021, 562,940 individuals (48.9% male) were present in the dataset with a mean age of 33.4 years (standard deviation: 18.4 years). The most frequently recorded diagnoses were substance use disorders (152,790 patients), major depressive disorder (129,120 patients) and anxiety disorders (103,923 patients). The median duration of follow-up was 7 months (interquartile range: 1.3 to 24.4 months).

Findings to date:

The dataset has supported epidemiological studies demonstrating increased risk of psychiatric hospitalization and reduced antidepressant treatment effectiveness among people with comorbid substance use disorders. It has also been used to develop data visualization 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 dataset will be further analyzed 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 health care services.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare real-world data analytics

Article Summary

Strengths and limitations

- The NeuroBlu dataset benefits from a large sample size of de-identified electronic health record (EHR) data from over 560,000 people who have received mental healthcare over a period of 21 years.
- The dataset is built on a robust de-identification pipeline and encryption framework that enables a wide range of users to safely analyze data through a trusted research environment using a graphical user interface or advanced analytic software (R and Python).
- Structured data on clinical severity (CGI-S) are recorded for over 80% of patients in the dataset. Natural language processing (NLP) enables access to rich clinical data from unstructured free text data as part of the Mental State Examination (MSE) and social factors that are not typically available in purely structured clinical datasets such as claims or RCT data.
- As the dataset 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.
- 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.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 U.S. 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 expectancy[3] and the lifetime costs associated with SMI are estimated to be $1.85 million 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 datasets, known as real world data (RWD), has improved the understanding of a range of medical disorders at a population level[7] and contributed to the development[8] and evaluation[9] of novel therapeutics. RWD may be analyzed to generate real-world evidence (RWE) to quantify the safety and effectiveness of therapeutic interventions. The generation of RWE from the analysis of largescale datasets that are representative of healthcare service delivery in routine settings provides a complementary source of data – as compared to more controlled clinical trials – that 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 characterize health disorders and to develop and deliver more effective treatments.[12] 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.[13] First, mental healthcare data are typically based on clinical evaluation of patients limited to diagnostic codes and subjective, non-standardized 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.[14] Third, most clinical data in mental healthcare EHRs are stored as unstructured free text that requires recoding into structured data prior to being analyzed. Relatedly, there is little standardization 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 de-identified mental health EHR datasets have been curated to support largescale population research,[15–17] provide clinical phenotype data for genomics research,[18] and to develop risk prediction tools with the potential to support clinical decision making.[19–22] 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,[23] environmental factors,[24] and clinical sentiment.[25] However, many

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

existing mental health EHR datasets are only accessible by clinical researchers based in specialized academic health science centers.[26] Enabling wider access, while maintaining robust data privacy and security standards, could substantially increase the impact of de-identified EHR datasets on RWE generation to improve clinical outcomes in people with mental disorders.

This paper presents a cohort profile of the NeuroBlu platform,[27] a trusted research environment which enables safe and secure analysis of de-identified EHR data from U.S. mental healthcare providers, including state-of-the-art NLP software to characterize mental state information from semi-structured text,[28] and a graphical user interface to enable users from a wide range of backgrounds to perform statistical analysis to support mental health RWE generation.

For peer review only

Cohort Description

Participants and setting

The NeuroBlu platform enables users to analyze de-identified data from over 560,000 patients receiving care from 25 U.S. mental healthcare providers that operate the MindLinc EHR system. Supplementary eTable 1 provides details of the locations, types of mental healthcare service and numbers of patients for each provider. In summary, the platform includes de-identified EHR data from individuals receiving mental healthcare from outpatient, inpatient, telemedicine, and residential care facilities in 12 U.S. 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.[29] MindLinc includes structured fields to record sociodemographic data, diagnoses, medications and clinical outcome scales as well as semi-structured free text fields to document the mental state examination (MSE) and treatment plan.[28] A subset of de-identified 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 de-identified at source before being transformed and normalized in a cloud-based U.S. Amazon Web Services (AWS) data warehouse and mapped to a common data model to create a harmonized dataset including data from all participating healthcare providers (eFigure 1). NeuroBlu users analyze the harmonized dataset 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.[30] In this way, de-identified EHR data may be analyzed by anyone with access to the trusted research environment without requiring any movement or disclosure of the underlying data.

De-identification procedure

MindLinc EHR data are de-identified at source (i.e. within the computing infrastructure of each participating healthcare provider) using the Safe Harbor method (eFigure 1, step 1).[31] 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 dataset. 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, URLs, social security numbers, 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 randomizing identifiers, recording age in years, and collecting the first three digits of zip codes with a population greater than 20,000).

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Transfer of de-identified EHR data from healthcare provider to secure AWS cloud

The de-identified EHR data are compressed and transferred from the source healthcare provider through a virtual private network (VPN) to a secure file transfer protocol (FTP) site (eFigure 1, step 2). The data are then transferred to a secure Amazon Elastic Compute Cloud (AWS EC2) instance (eFigure 1, step 3) where they are decompressed and assembled for loading into a PostgreSQL database (eFigure 1, step 4).

Given the relational nature of the MindLinc databases, the files produced by the EHR data de-identification 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 de-identification process is irreversible, preventing re-identification 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 (eFigure 1, step 4) and is further described in section 1 of the Supplementary 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 (e.g. appearance and behavior) or may be elicited through direct questioning or the presentation of cognitive tasks. The MSE provides a rich source of clinical phenotype data which characterizes the nature of presenting symptoms which could be associated with varying clinical outcomes and response to treatment.

The MindLinc EHR includes a semi-structured "status assessment" field in which clinicians can document features associated with a patient's MSE. The status assessment field allows clinicians to choose pre-defined features from a list of options and/or to document findings as unstructured free text. However, pre-defined 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 pre-defined features in MindLinc is 0% whereas the median percentage with unstructured free text is 67%.

NLP is a text mining technique which enables automated extraction and classification of features from unstructured free text. 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.[28] 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 de-identified EHR dataset to create a table of NLP-derived MSE measures which may be joined and analyzed with other structured clinical and sociodemographic data in the NeuroBlu dataset. 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 below in the section on *Unstructured data*.

Data variables

Full details of available data variables and their structure within the NeuroBlu MindLinc dataset are provided in section 2 of the Supplementary Material. The data variable structure is derived from the structure of the source MindLinc EHR and transformed into a relational database which may be

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

accessed using predefined database queries in the front-end interface or through embedded SQL queries in R scripts (described subsequently).

Structured data

The dataset includes structured sociodemographic data on gender, year of birth, race, ethnicity, U.S. 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 dataset that include data on contacts with mental health services, emergency room visits, hospital admissions, 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),[32] Clinical Global Impression - Improvement (CGI-I),[32] Global Assessment of Functioning (GAF)[33] and Montgomery-Asberg Depression Rating Scale (MADRS).[34]

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 dataset. Structured questionnaire data are available for the following numbers of patients within the dataset:

(i) CGI-S: 471,256 patients (83.7%)
(ii) CGI-I: 361,819 patients (64.3%)
(iii) GAF: 310,895 patients (55.3%)
(iv) MADRS: 1,126 patients (0.2%)

The dataset 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 Supplementary 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 Supplementary Material.

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.[35]

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; Judgment; Language; Level of consciousness; Memory; Mood; Orientation; Psychomotor; Reasoning;

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Sensorium; Sleep; Speech; Suicidal; 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 (e.g. 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 analyzed 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 to patients with prominent positive psychotic symptoms.[36] 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 target 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: (i) Cohort Builder, (ii) Category Mapper, (iii) Data Explorer and (iv) 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 dataset 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 hospitalized) 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 (eFigure 3). This enables users to pre-define 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 dataset is provided in the Data Explorer with accompanying Dataset Specifications that describe the data structure (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

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 dataset 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 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: (i) *NeuroBlu*: utility functions which simplify data assembly processes using prebuilt SQL queries, (ii) *SurvBlu*: a framework for conducting time-to-event analyses, and (iii) *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 visualizing data.

Data privacy and ethics

Under HIPAA regulations, U.S. healthcare providers record and store personal healthcare information for the purposes of providing treatment, billing data, to support healthcare service operations. De-identified data may be used for secondary analyses. The NeuroBlu platform has received a waiver of authorization for analysis of de-identified healthcare data from the WCG Institutional Review Board (Ref: WCG-IRB 1-1470336-1).

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 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 is encrypted with AES256.

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

Participant characteristics

Descriptive statistics for the NeuroBlu dataset were obtained on 31st July 2021. A total of 562,940 individuals (48.9% male) were present in the dataset with a mean age of 33.4 years (standard deviation: 18.4 years). The majority of the population was White (49.1%). eTable 2 and eFigure 5

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

provide a full breakdown of the population by race. Children and adolescents up to 19 years of age account for 28.7% of the dataset, working age adults (aged 20 to 59 years) account for 62.2% and adults aged 60 years or above account for 9.1% (eTable 3/eFigure 6).

eTable 4 and Figure 3 illustrate the prevalence of mental disorders in the dataset. 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 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 & Alzheimer's disease, bipolar disorder, major depressive disorder and anxiety disorders among female patients. The constituent diagnostic codes are provided in eTable 6.

eTable 7 and eFigure 8 provide a breakdown by U.S. State of Residence. The most frequent U.S. 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 eTable 8 and eFigure 9. The greatest mean CGI-S scores were found in people with schizoaffective disorder (4.48, standard deviation: 1.42) and schizophrenia (4.44, std dev: 1.44) with the lowest mean score recorded in people with substance-use disorders (3.61, std dev: 1.80). A further analysis was conducted among patients with at least 2 recorded CGI-S scores to estimate the mean maximum and mean minimum score by diagnosis (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 characterized by periods of remission in between episodes.

The number of patients with at least one hospital visit (inpatient or emergency department) by diagnosis (eTable 10/eFigure 10) 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 hospitalization compared to 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 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 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.

eFigure 11 provides a distribution of the total number of clinical encounters (visits) per patient. Patients had a median of 5 visits recorded (interquartile range: 2 to 20 visits). eFigure 12 provides a

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

distribution of the total follow-up duration available for each patient in the dataset. The median follow-up duration was 7 months (IQR: 1.3 to 24.4 months).

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

(i) Gender: 1,572 patients (0.23%)
(ii) Race: 186,517 patients (33.1%)
(iii) Psychiatric diagnosis: 66,064 patients (11.7%)
(iv) State: 21,919 patients (3.9%)
(v) NLP-derived MSE data: 132,690 patients (23.6%)

Patient and Public Involvement

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

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Findings to date

The NeuroBlu platform and MindLinc EHR dataset has been used to generate real world evidence through largescale mental health epidemiology studies as well as to develop automated NLP, data visualization and predictive analytic tools based on real world EHR data.

Epidemiology of mental disorders

Data from people with substance use disorders has been analyzed to investigate its impact among minority ethnic people,[37] and demonstrated increased rates of psychiatric hospitalization among children and adolescents[38] and increased risk of psychiatric comorbidities and hospitalization among adults with substance use disorders.[39]

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.[40] Augmentation of antidepressant treatment with a second generation antipsychotic in people with depression was more likely in those with a high baseline CGI-S.[41]

Longitudinal prescribing data has been analyzed to demonstrate that antidepressant prescribing reduced in response to U.S. Food and Drug Administration (FDA) warnings on risk of suicidality in children and adolescents.[42] 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.[43]

Mental health informatics

As well as being analyzed to generated real world evidence, the dataset has been used to develop novel data visualization and predictive analytic methods to support the analysis of mental health real world data including a data visualization tool to support clinical decision making,[44] a model to predict treatment response in major depressive disorder,[45] infrastructure to augment data from clinical trials with clinician-administered rating scales from the EHR,[46] and a data visualization tool to evaluate comparative effectiveness of psychotropic medications.[47]

Furthermore, the availability of unstructured free text data has supported the development of NLP tools to automatically extract clinically meaningful information from the dataset, as described previously.[28,48]

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Strengths and limitations

The NeuroBlu dataset contains a rich source of de-identified EHR data that supports the generation of real-world evidence to improve understanding of mental disorders and factors affecting clinical outcomes. NeuroBlu includes data on over 560,000 people attending 25 U.S. mental healthcare providers over a period of 21 years. The large sample size and long duration of follow-up enables 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 randomized controlled trials (RCTs) or prospective observational studies (for example, comparative effectiveness studies on multiple medications or assessing the impact of substance use disorders on clinical outcomes).

NeuroBlu is built upon a robust de-identification pipeline that ensures no personal identifiable data are available to users who analyze the data. A strong data security framework that employs internationally validated encryption standards protects the anonymized dataset from being accessed outside the trusted research environment of NeuroBlu. These safeguards ensure the security of the dataset while still enabling users to easily analyze 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 visualize 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 analyze the data while ensuring the data remain secure within the trusted research environment. These features allow users from a wide range of backgrounds to analyze the data without requiring expertise in statistical software while at the same time enabling expert data scientists to perform advanced analytical procedures on the dataset using R or Python.

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

The dataset 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 dataset 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 analyzing EHR datasets containing only structured data or claims data.

As EHR data are recorded as part of routine clinical practice, a limitation of the dataset is that clinical data may not be comprehensively completed across all domains for all patients. Clinicians typically record data which 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

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

(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.

The dataset 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 dataset 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 (e.g. acute physical healthcare providers or diagnoses and medications recorded in community primary care settings) are not available within the NeuroBlu dataset.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Future plans

As patients continue to receive care from healthcare providers using the MindLinc EHR, the NeuroBlu dataset will grow over time increasing in sample size and follow-up duration. We plan to analyze these data to generate real-world evidence to better understand the factors related to poor mental health outcomes, to evaluate the impact of treatments, and to develop predictive analytic tools that quantify these factors at individual patient level.[49] 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 personalizing 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 enables the MindLinc dataset to be analyzed by a wide range of users from varied professional disciplines in any geographical location. To this end, we highly encourage collaboration with individuals, academic institutes, healthcare providers and commercial organizations who wish to analyze the dataset 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 de-identification pipeline and trusted research environment to support real world evidence generation from their own datasets. 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.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare real-world data analytics

Figure captions

Figure 1: NeuroBlu Cohort Builder illustrating the population characteristics of female patients diagnosed with Post Traumatic Stress Disorder who have received sertraline

Figure 2: NeuroBlu R Code Engine which 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

Figure 3: Prevalence of mental disorders in the NeuroBlu dataset

Figure 4: Mean maximum and mean minimum CGI-S score by diagnosis

Figure 5: Percentage of patients with at least one NLP-derived MSE feature by diagnosis

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Further details

Competing interests: All authors have completed the ICMJE uniform disclosure form and declare: 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.

Author contributions: 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, MR, JS and SK) contributed to cohort profile design, manuscript preparation and approved the final version.

Ethics approval: The NeuroBlu platform has received a waiver of authorization for analysis of de-identified healthcare data from the WCG Institutional Review Board (Ref: WCG-IRB 1-1470336-1).

Source of funding: The NeuroBlu dataset and MindLinc EHR system are owned and funded in full by Holmusk.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare real-world data analytics

References

1 Rehm J, Shield KD. Global Burden of Disease and the Impact of Mental and Addictive Disorders. *Curr Psychiatry Rep* 2019;**21**:10. doi:10.1007/s11920-019-0997-0

2 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. https://www.samhsa.gov/data/sites/default/files/reports/rpt29393/2019NSDUHFFRPDFWHTML/2019NSDUHFFR1PDFW090120.pdf

3 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.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1563985/

4 Seabury SA, Axeen 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. doi:10.1377/hlthaff.2018.05246

5 Degenhardt L, Charlson F, Ferrari A, *et al.* 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. *The Lancet Psychiatry* 2018;**5**:987–1012. doi:10.1016/S2215-0366(18)30337-7

6 Mental Health America. 2020 Access to Care Data. Ment. Heal. Am. 2020.https://www.mhanational.org/issues/2020/mental-health-america-access-care-data#four

7 Maissenhaelter BE, Woolmore AL, Schlag PM. Real-world evidence research based on big data: Motivation—challenges—success factors. *Onkologe* 2018;**24**:91–8. doi:10.1007/s00761-018-0358-3

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. doi:10.1016/j.drudis.2017.12.002

9 Corrigan-Curay J, Sacks L, Woodcock J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. *JAMA - J Am Med Assoc* 2018;**320**:867–8. doi:10.1001/jama.2018.10136

10 Klonoff DC. The New FDA Real-World Evidence Program to Support Development of Drugs and Biologics. *J Diabetes Sci Technol* 2020;**14**:345–9. doi:10.1177/1932296819832661

11 Roberts MH, Ferguson GT. Real-World Evidence: Bridging Gaps in Evidence to Guide Payer Decisions. *PharmacoEconomics - Open* 2021;**5**:3–11. doi:10.1007/s41669-020-00221-y

12 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–e1912869. doi:10.1001/jamanetworkopen.2019.12869

13 Stewart R, Davis K. 'Big data' in mental health research: current status and emerging possibilities. *Soc Psychiatry Psychiatr Epidemiol* 2016;**51**:1055–72. doi:10.1007/s00127-016-1266-8

14 Peterson BS. Editorial: Biomarkers in precision medicine for mental illnesses. *J Child Psychol Psychiatry Allied Discip* 2020;**61**:1279–81. doi:10.1111/jcpp.13357

15 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**. doi:10.1136/bmjopen-2015-008721

16 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. doi:10.1371/journal.pone.0190703

17 Cardinal RN. Clinical records anonymisation and text extraction (CRATE): An open-source software system. *BMC Med Inform Decis Mak* 2017;**17**:50. doi:10.1186/s12911-017-0437-1

18 Smoller JW. The use of electronic health records for psychiatric phenotyping and genomics. *Am J Med Genet Part B Neuropsychiatr Genet* 2018;**177**:601–12. doi:10.1002/ajmg.b.32548

19 Holderness E, Miller N, Cawkwell P, *et al.* Analysis of risk factor domains in psychosis patient health records. *J Biomed Semantics* 2019;**10**:19. doi:10.1186/s13326-019-0210-8

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

20 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* Published Online First: 2020. doi:https://doi.org/10.1016/j.schres.2020.09.016

21 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. doi:10.1016/j.schres.2020.05.007

22 Raket LL, Jaskolowski J, Kinon BJ, *et al.* Dynamic ElecTronic hEalth reCord deTection (DETECT) of individuals at risk of a first episode of psychosis: a case-control development and validation study. *Lancet Digit Heal* 2020;**2**:e229–39. doi:10.1016/S2589-7500(20)30024-8

23 Jackson RG, Patel R, Jayatilleke 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**. doi:10.1136/bmjopen-2016-012012

24 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. doi:10.1136/bmjopen-2020-042949

25 Holderness E, Cawkwell P, Bolton K, *et al.* 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. doi:10.18653/v1/w19-1915

26 Ford T, Mansfield KL, Markham S, *et al.* The challenges and opportunities of mental health data sharing in the UK. *Lancet Digit Heal* 2021;**3**:e333–6. doi:10.1016/s2589-7500(21)00078-9

27 Holmusk. NeuroBlu: The Evidence Footprint. Building the future of behavioral health with Real-World Data and analytics. 2021.https://www.neuroblu.ai/

28 Mukherjee SS, Yu J, Won Y, *et al.* Natural Language Processing-Based Quantication of the Mental State of Psychiatric Patients. *Comput Psychiatry* 2020;**4**:76. doi:10.1162/cpsy_a_00030

29 Gersing K, Krishnan R. Clinical Computing: Clinical Management Research Information System (CRIS). *Psychiatr Serv* 2003;**54**:1199–200. doi:10.1176/appi.ps.54.9.1199

30 Response to the Goldacre Review - “Trusted Research Environments and Data Management”. HDR UK. 2021.https://www.hdruk.ac.uk/news/response-to-the-goldacre-review-trusted-research-environments-and-data-management/

31 Kayaalp M. Modes of De-identification. *AMIA Annu Symp Proc* 2018;**2017**:1044–50.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977668/

32 Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. *Psychiatry (Edgmont)* 2007;**4**:28–37.https://pubmed.ncbi.nlm.nih.gov/20526405

33 *Diagnostic and statistical manual of mental disorders : DSM-IV*. Fourth edition. Washington, DC : American Psychiatric Association, [1994] ©1994 https://search.library.wisc.edu/catalog/999733358502121

34 Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry J Ment Sci* 1979;**134**:382–9. doi:10.1192/bjp.134.4.382

35 Allen J, Balfour R, Bell R, *et al.* Social determinants of mental health. *Int Rev psychiatry* 2014;**26**:392–407.

36 Carbon M, Correll CU. Thinking and acting beyond the positive: The role of the cognitive and negative symptoms in schizophrenia. *CNS Spectr* 2014;**19**:38–52. doi:10.1017/S1092852914000601

37 Wu LT, 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. doi:10.1016/j.jpsychires.2013.08.022

38 Wu LT, 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*

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Psychiatr Res 2011;**45**:1453–62. doi:10.1016/j.jpsychires.2011.06.012

39 Wu LT, 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. doi:10.1016/j.jpsychires.2012.12.009

40 Lee TS, Assam PN, Gersing KR, *et al.* The effectiveness of antidepressant monotherapy in a naturalistic outpatient setting. *Prim Care Companion J Clin Psychiatry* 2012;**14**. doi:10.4088/PCC.12m01364

41 Gersing KR, Sheehan JJ, Burchett B, *et al.* Use of augmentation agents for treating depression: Analysis of a psychiatric electronicmedical record data set. *Psychiatr Serv* 2014;**65**:1062–5. doi:10.1176/appi.ps.201300288

42 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. doi:10.1007/s11126-015-9389-8

43 Anisha Balani Sheryl Ker, Jennifer Hsu, Dustin Demoss, Miguel E. Rentería Sankha S. Mukherjee AJR. T4. Buprenorphine treatment retention in individuals with opioid use disorder: insights from real-world dataset. In: *American Society of Clinical Psychopharmacology*. 2021. https://pmg.joynadmin.org/documents/1005/60ad2955ef0d26cf5e684539.pdf#page=181

44 Mane KK, Owen P, Schmitt C, *et al.* 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. doi:10.1145/2506583.2506608

45 Akushevich I, Kravchenko J, Gersing K, *et al.* 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. doi:10.1145/2506583.2506618

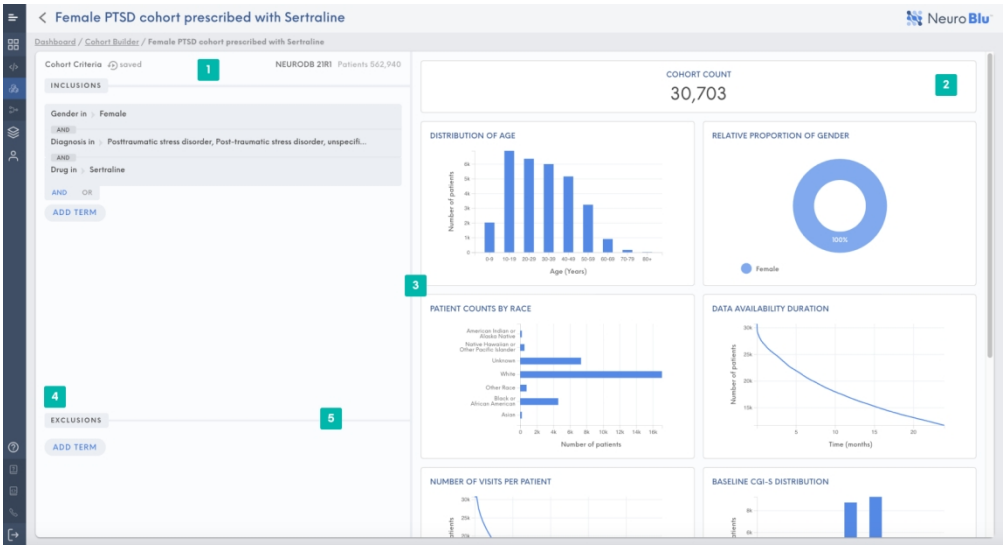
46 Pan JJ, Nahm M, Wakim P, *et al.* A centralized informatics infrastructure for the National Institute on Drug Abuse Clinical Trials Network. *Clin Trials* 2009;**6**:67–75. doi:10.1177/1740774508100983

47 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. doi:10.1016/j.jbi.2011.09.003

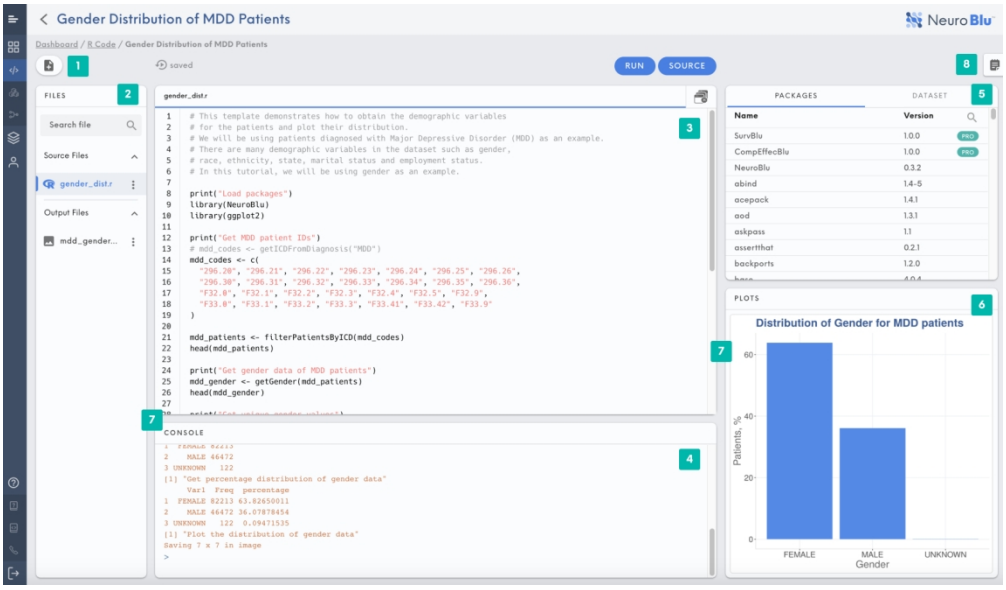
48 Sankha Mukherjee Yida Won, Mary McClay, Lu Wang, John Rush, Joydeep Sarkar JW. W35. Natural language processing-based quantification of the mental state of psychiatric patients. In: *American Society of Clinical Psychopharmacology*. 2021. https://pmg.joynadmin.org/documents/1005/60ad2955ef0d26cf5e684539.pdf#page=151

49 Hahn T, Nierenberg AA, Whitfield-Gabrieli S. Predictive analytics in mental health: applications, guidelines, challenges and perspectives. *Mol Psychiatry* 2017;**22**:37–43. doi:10.1038/mp.2016.201

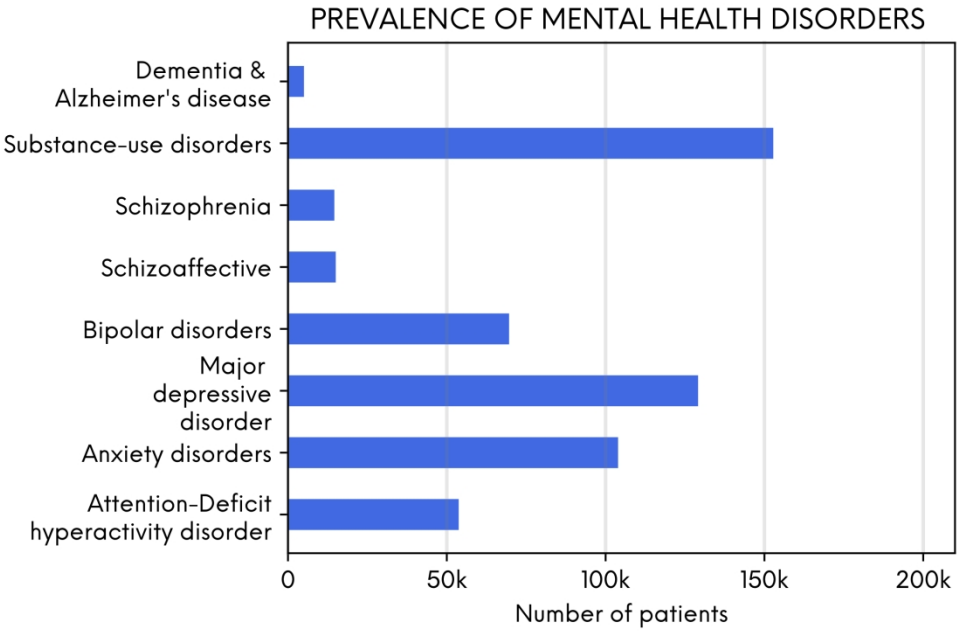
For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml



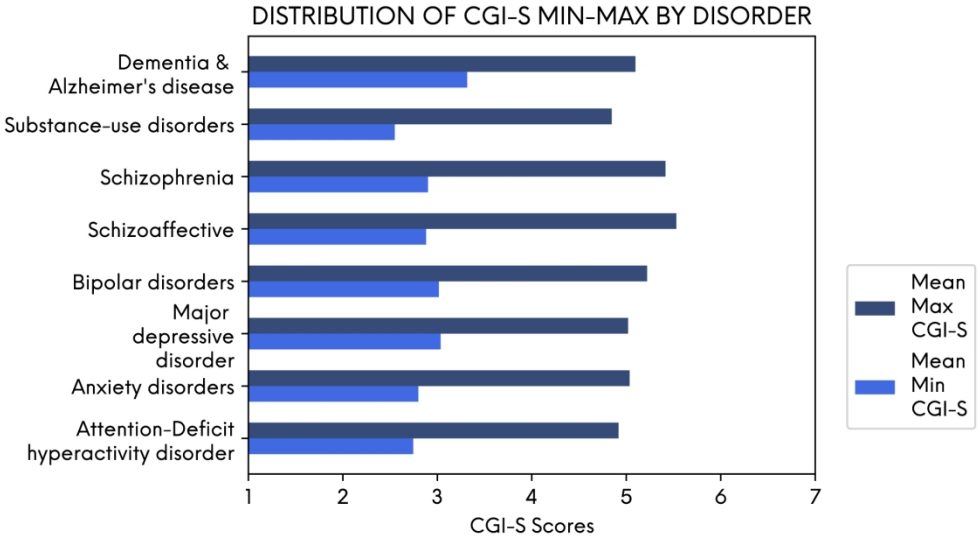
NeuroBlu Cohort Builder illustrating the population characteristics of female patients diagnosed with Post Traumatic Stress Disorder who have received sertraline



NeuroBlu R Code Engine which 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

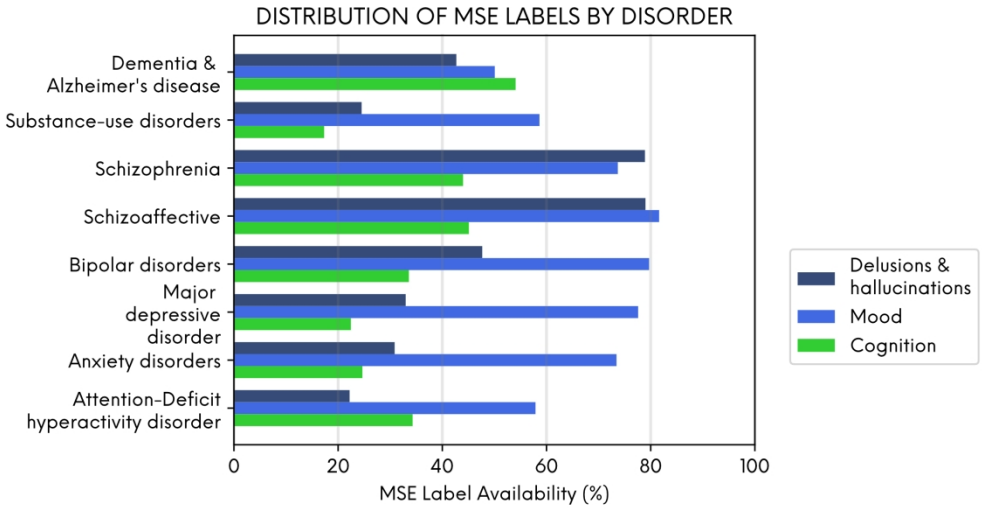


Prevalence of mental disorders in the NeuroBlu dataset



Mean maximum and mean minimum CGI-S score by diagnosis

177x99mm (300 x 300 DPI)



Percentage of patients with at least one NLP-derived MSE feature by diagnosis

Online supplementary material

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare real-world data analytics

Table of contents

Table of contents 1
1. NeuroBlu Common Data Model (CDM) pipeline 3
(i) Structural Normalization 3
(ii) Semantic Normalization 4
(iii) Data Validation 4
2. Data variables 5
(i) Clinical Data Tables 5
person 6
observation_period 8
visit_occurrence 9
condition_occurrence 10
drug_exposure 11
measurement 12
observation 14
(ii) Family History 17
(iii) Derived Measurements 18
(iv) Mental State Examination 23
eTable 1: Description of U.S. clinics contributing EHR data to NeuroBlu 31
eTable 2: NeuroBlu dataset race distribution 32
eTable 3: NeuroBlu dataset age distribution 32
eTable 4: NeuroBlu dataset diagnosis distribution 32
eTable 5: Gender distribution by diagnosis 33
eTable 6: ICD-9/ICD-10 code definitions for diagnostic category 34
eTable 7: U.S. State of Residence 65
eTable 8: Mean CGI-S score by diagnosis 67
eTable 9: Mean maximum and minimum CGI-S score by diagnosis 68
eTable 10: Number of patients with at least one hospital visit by diagnosis 69
eTable 11: SAV features included in MSE NLP analysis (eTable 12) 70
eTable 12: Number of patients with at least one MSE feature by diagnosis 71
eFigure 1: MindLinc EHR data de-identification and harmonization pipeline 72

For peer review only

Online supplementary material

eFigure 2: OMOP CDM schematic 73

eFigure 3: NeuroBlu Category Mapper illustrating the classification of individual drugs into different types of medication 74

eFigure 4: NeuroBlu Data Explorer which provides a summary of the dataset and allows the top 50 rows of each table to be viewed 75

eFigure 5: NeuroBlu dataset race distribution 76

eFigure 6: NeuroBlu dataset age distribution 77

eFigure 7: Gender distribution by diagnosis 78

eFigure 8: U.S. State of Residence 79

eFigure 9: Mean CGI-S score by diagnosis 80

eFigure 10: Percentage of patients with at least one hospital visit by diagnosis 81

eFigure 11: Distribution of number of visits per patient 82

eFigure 12: Distribution of duration of follow-up 83

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

1. NeuroBlu Common Data Model (CDM) pipeline

The NeuroBlu Data Model is based on the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) developed by the Observation Health Data Sciences and Informatics (OHDSI) community (eFigure 1). The model enables data contained within disparate observational databases to be transformed into a common format (the "data model") as well as a common representation (terminologies, vocabularies, coding schemes). This allows systematic analyses to be performed using a library of standard analytical routines that have been written based on the OMOP CDM.

The NeuroBlu Data Model includes the following modifications of the OMOP CDM to better support the representation of mental healthcare data:

1. Representing ICD9/ICD10 diagnostic codes in the CONDITION_OCCURRENCE table rather than the specified SNOMED concepts, as users typically have more familiarity with the ICD code sets than the SNOMED code set.
2. Adding the current value of mutable demographic attributes such as marital status, education, and occupation to the PERSON to allow users to access this information without needing to join a separate table.
3. Adding custom tables to support derived data and other data elements that do not have a natural home in the OMOP CDM. This includes derived measurements like smoothened CGI-S scores, imputed MADRS scores, and Mental State Examination (MSE) data obtained through natural language processing (NLP).

Once the MindLinc data are de-identified, transferred, and loaded into the relational PostgreSQL database, the remaining data integration processes required to ensure the data are ready for analysis within the NeuroBlu dataset include structural normalization, semantic normalization, and data validation.

(i) Structural Normalization

Structural normalization is the process of transforming the data into a single schema or layout. For example, to examine patient demographic information in a consistent way, regardless of how it was stored in the originating EHRs, the various patient and member demographic data are organized into a single, commonly structured PERSON table. This requires establishing and implementing various transformation rules that get applied to the data, such as those described below:

1. **Source-to-Target Mapping Logic:** this refers to the conveyance of information from a specific place in the source data to its related place in the target data model. For example, the MindLinc EHR source data contains a BACKGROUND table to record patient demographic attributes, such as RACE and SEX. In order for these attributes to end up in the appropriate place in the NeuroBlu Data Model, they need to be mapped from the relevant source field to the associated target field. In this example, the information in RACE and SEX MindLinc database would end up in the RACE_SOURCE_VALUE and GENDER_SOURCE_VALUE fields of the PERSON table in the NeuroBlu database.
2. **Inclusion/Exclusion Logic:** this refers to the criteria applied to the data in order for them to be included in the NeuroBlu Data Model. For example, information from the MindLinc source data that is marked as "entered in error" or "erroneous" brings into question whether it is truly evidence from real world data and in turn is excluded from the NeuroBlu Data Model.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

3. **De-duplication Logic:** this refers to the removal of duplicative or outdated information, such that only up-to-date data are pulled into the NeuroBlu Data Model and are only pulled in once. Depending on the source data elements of interest and the mechanism by which updates are applied to them, the method for de-duplication will vary. In the case of the MindLinc data, a destructive refresh approach is taken across all data tables such that the most recently acquired data will always replace the older data.

(ii) Semantic Normalization

Semantic normalization is the process of translating the data to a single set of units and vocabulary. For example, in order to specifically examine female patients in the aforementioned PERSON table, the various ways the patient and member demographic data codify the concept of "female" need to be examined such that a translation between the source codes (e.g. "F", "FEM", "FEMALE", "0", "396123", etc.) and the NeuroBlu standard concept ("FEMALE") can be created and applied to the data. These translations are generated through a combination of direct-mapping, text-string-matching, and NLP based on a thorough examination of the native codes, their descriptions, any supporting information provided about them (e.g. reference ranges for scores, generic name for medications, etc.), and, if necessary, expert input from the healthcare provider organization they are originating from.

A number of industry standard reference code sets are leveraged as part of the semantic normalization process for the MindLinc data, including RxNorm for drug mapping and ICD9/ICD10 for diagnostic code mapping. The target concept sets for all normalized data elements are the Standardized Vocabulary provided as part of the OMOP CDM standard.

(iii) Data Validation

A core component of data acquisition and integration is ensuring that the resultant data set is of the highest quality possible, where quality is defined as a function of conformance, completeness, correctness, and consistency. In order to measure the quality of newly integrated data against these key dimensions, the NeuroBlu CDM pipeline employs an approach to data validation that uses the wealth of data available during the process to produce a composite measure of what certain key data points typically look like in different situations, and then evaluates the data against these expectations. Doing this efficiently requires validation of the data as they flow through the processing pipeline, which in turn necessitates a robust strategy for determining key quality indication metrics, generating those metrics in a systematic manner, and distilling those metrics into meaningful insights.

The Data Validation process works in concert with the above normalization processes to identify issues or inconsistencies in the data that need to be addressed and provide guidelines on how those various issues or inconsistencies should be handled. Erroneous data that can't be fixed at the source are either remedied or removed by NeuroBlu's Extract, Transform, Load (ETL) processes, while quirks or unexpected features of the data that are not necessarily errors are systematically logged.

The end result of these data integration processes is the creation of a single normalized and validated Common Data Model for use in the NeuroBlu trusted research environment.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

2. Data variables

(i) Clinical Data Tables

The clinical data tables contain the core information about the clinical events that occurred longitudinally during valid Observation Periods for each person, as well as demographic information for the person. In addition to the standardized clinical data tables defined in the OMOP CDM, the NeuroBlu dataset also contains other tables that are not defined in the OMOP CDM including data on Family History, Derived Measurements and Mental State Examination and which are described in greater detail in sections (ii), (iii) and (iv). The tables form part of a relational database in which "_id" columns may be joined between tables to assemble clinical data at individual person level.

Table name	Type	Content
person	CDM	Socio-demographic information of each patient including year of birth, gender, race, location of residence, occupational status and education status.
observation_period	CDM	Time interval for which a patient has data recorded. Data from this table is used to filter cohorts according to minimum duration within the dataset to support longitudinal analyses.
visit_occurrence	CDM	Records of each visit and its type. Data may be obtained from this table on the number of contacts with the treating mental healthcare service and the type of contact (e.g. outpatient or inpatient).
condition_occurrence	CDM	Records of clinical diagnosis in the form of ICD-9-CM and ICD-10-CM codes. Data from this table is used to generate population cohorts based on primary mental disorder diagnosis and to obtain data on comorbid diagnoses.
drug_exposure	CDM	Records of any exposure to prescribed medications including the names of prescribed medications, start and stop dates, dose and frequency of administration.
measurement	CDM	Records of direct clinical measurements using clinician-rated structured rating scales.
observation	CDM	NLP-derived records of social factors which could contribute towards risk of developing a mental disorder and/or poorer clinical outcomes. This table can also be used to obtain social factor data as outcomes (e.g. social and occupational functioning)
family_history	Non-CDM	Provides recorded data on family history of mental disorders
derived_measurement	Non-CDM	Clinical measurements which have been transformed or derived from recorded CGI-S scores.
derived_mse	Non-CDM	NLP-derived mental state examination (MSE) data.
CDM: OMOP Common Data Model Non-CDM: NeuroBlu customized data structure (not defined in OMOP)		

For peer review only

Online supplementary material

person

The person table contains records that uniquely identify each person in the dataset. This table also contains demographic information about each individual (excluding data removed by the de-identification process). All other clinical data tables refer to this table using the person_id field which is the unique record identifier for each individual represented in the dataset.

Columns

Field	Type	Description
person_id	int8	A unique identifier for each person. This value is used to join onto other tables in NeuroBlu to obtain clinical data at individual person level.
gender_concept_id	int4	A foreign key that refers to an identifier in the concept table for the unique gender of the person, belonging to the 'Gender' Vocabulary. Gender data are obtained by joining on this value (see below).
year_of_birth	int4	The year of birth of the person.
race_concept_id	int4	A foreign key that refers to an identifier in the concept table for the unique race of the person, belonging to the 'Race' Vocabulary. Race data are obtained by joining on this value (see below).
ethnicity_concept_id	int4	A foreign key that refers to an identifier in the concept table for the unique ethnicity of the person, belonging to the 'Ethnicity' Vocabulary. Ethnicity data are obtained by joining on this value (see below).
zip_code_source_value	varchar	3-digit ZIP code of the person's location.
city	varchar	City name based on 3-digit ZIP code.
state	varchar	State name based on 3-digit ZIP code.
marital_status	varchar	Marital status of the person.
occupation	varchar	Employment status of the person.
education	int2	Number of years of formal education. When education_source_value is not available, defaults to 0
education_source_value	varchar	Source value for education. Not always available.

Conventions

No.	Convention Description
1	All tables representing patient-related domains have a foreign-key reference to the person_id field in the person table.
2	Each person record has associated demographic attributes which are assumed to be constant for the patient throughout the course of their periods of observation. For example, the location or gender is expected to have a unique value per person, even though in life these data may change over time.
3	The month and day of birth for a person is not recorded. Therefore, value defaults to Jan 1st.
4	Ethnicity and Race in the OMOP CDM follow the 1997 Office of Management and Budget (OMB) standards for data on race and ethnicity.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Gender Vocabulary

concept_id	concept_name
8507	MALE
8532	FEMALE
8551	UNKNOWN

Race Vocabulary

concept_id	concept_name
8515	Asian
8516	Black or African American
8522	Other Race
8527	White
8552	Unknown
8557	Native Hawaiian or Other Pacific Islander
8657	American Indian or Alaska Native

Ethnicity Vocabulary

concept_id	concept_name
759814	Ethnic group unknown
38003563	Hispanic or Latino
38003564	Not Hispanic or Latino

Marital Status Vocabulary

marital_status
death of spouse
divorced
engaged
married
separated
single
together
unknown

Occupation Vocabulary

occupation
Disabled
Employed
Retired
Student
Unemployed
Unknown

For peer review only

observation_period

The observation_period table contains records which uniquely define the spans of time for which a patient is expected to have clinical events recorded within the source systems, even if no events in fact are recorded (healthy patient with no healthcare interactions).

Columns

Field	Type	Description
observation_period_id	serial	A unique identifier for each observation period.
person_id	int8	A foreign key identifier to the person for whom the observation period is defined. The demographic details of that person are stored in the person table.
observation_period_start_date	date	The start date of the observation period for which data are available from the data source.
observation_period_end_date	date	The end date of the observation period for which data are available from the data source.
period_type_concept_id	int4	A foreign key identifier to the predefined Concept in the Standardized Vocabularies reflecting the source of the observation period information, belonging to the 'Obs Period Type' Vocabulary.

Conventions

No.	Convention Description
1	Each person has to have at least one observation period.
2	One person may have one or more disjoint observation periods, during which times analyses may assume that clinical events would be captured if observed.
3	Each person can have more than one valid observation period record, but no two observation periods can overlap in time for a given person.
4	As a general assumption, during an observation period any clinical event that happens to the patient is expected to be recorded. Conversely, the absence of data indicates that no clinical events occurred to the patient.
5	Both the start date and the end date of the clinical event has to be between observation_period_start_date and observation_period_end_date.
6	For EHR data, the observation period cannot be determined explicitly, because patients usually do not announce their departure from a certain healthcare provider. The extract, transform, load (ETL) process will have to apply some heuristic to make a reasonable estimate on what the observation_period should be.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

visit_occurrence

The visit_occurrence table contains the spans of time a patient continuously receives medical services from one or more providers at a care site in a given setting within the health care system.

Columns

Field	Type	Description
visit_occurrence_id	int8	A unique identifier for each person's visit or encounter at a healthcare provider.
person_id	int8	A foreign key identifier to the person for whom the visit is recorded. The demographic details of that person are stored in the person table.
visit_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies belonging to the 'Visit' Domain.
visit_start_datetime	timestamp	The date and time when the visit started. The time is not captured in the source and defaults to 00:00:00.
visit_end_datetime	timestamp	The date and time when the visit ended. Not recorded in the source. Takes the same value as visit_start_datetime.
visit_type_concept_id	int4	A foreign key identifier to the predefined Concept in the Standardized Vocabularies reflecting the source data from which the visit record is derived from, belonging to the 'Visit Type' Vocabulary. Visit type data are obtained by joining on this value (see below).
provider_id	int8	A foreign key to the primary care provider the person is seeing.
care_site_id	int8	A foreign key to the site of primary care.
visit_source_value	varchar	Source value for visit_concept_id.

Conventions

No.	Convention Description
1	A visit occurrence is recorded for each engagement of a patient with the healthcare system. It reflects any interaction with the healthcare service, not just in person clinical interactions.
2	Valid Visit Concepts belong to the 'Visit' domain.
3	At any one day, there could be more than one visit.
4	Visit times and visit end dates are not captured in the source. They resort to 00:00:00 and visit start dates, respectively.

Visit types

The table below describes the different visit types documented in the visit_occurrence table.

concept_id	concept_name
262	Emergency Room and Inpatient Visit
9201	Inpatient Visit
9202	Outpatient Visit
32693	Health examination
581458	Pharmacy visit
581477	Office Visit
42898160	Non-hospital institution Visit

For peer review only

Online supplementary material

condition_occurrence

The condition_occurrence table contains the presence of a disease or medical condition recorded as a diagnosis according to the International Classification of Disease (ICD), which is either observed by a provider or reported by the patient. In the NeuroBlu dataset, diagnoses are recorded as ICD-9-CM and ICD-10-CM codes in the source.

Columns

Field	Type	Description
condition_occurrence_id	serial	A unique identifier for each condition occurrence event.
person_id	int8	A foreign key identifier to the person for whom the condition is recorded. The demographic details of that person are stored in the person table.
condition_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies belonging to the 'Condition' Domain. Joining on this value extracts the corresponding ICD-9-CM or ICD-10-CM diagnostic code.
condition_start_datetime	timestamp	The date and time when the instance of the condition is recorded. The time is not captured in the source and defaults to 00:00:00.
provider_id	int8	A foreign key to the primary care provider the person is seeing.
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the condition was diagnosed.
condition_source_value	varchar(255)	Source value for condition_concept_id.

Conventions

No.	Convention Description
1	Past diagnoses ('history of') are not recorded in the condition_occurrence table. Instead, they are listed in the observation table.
2	The exact start date and end date for a condition is not recorded explicitly. Rather, the start date contains the date when the diagnosis was made, and the end date is left empty.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

drug_exposure

The drug_exposure table contains records about the utilization of a prescribed medication when ingested or introduced to the body. Prescribed medications are inferred from clinical events associated with orders, prescriptions written, pharmacy dispensing, procedural administrations, and other patient-reported information.

Columns

Field	Type	Description
drug_exposure_id	serial	A unique identifier for each condition exposure event.
person_id	int8	A foreign key identifier to the person for who is subjected to the drug. The demographic details of that person are stored in the person table.
drug_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies belonging to the 'Drug' Domain. Joining on this value extracts the name of the prescribed medication.
drug_exposure_start_datetime	timestamp	The start date and time for the current instance of drug utilization. Valid entries include a start date of a prescription or the date on which a drug was administered. The time is not captured in the source and defaults to 00:00:00.
drug_exposure_end_datetime	timestamp	The end date and time for the current instance of drug utilization. It could be a known or an inferred date and denotes the last day at which the patient was still exposed to the drug during an instance of drug utilization.. The time is not captured in the source and defaults to 00:00:00.
quantity	numeric	Total quantity of drug exposed, in mg. Not always available.
days_supply	int4	The number of days of drug exposure. Not always available.
provider_id	int8	A foreign key to the primary care provider the person is seeing.
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the drug exposure was initiated.
drug_source_value	varchar(50)	Source value for drug_concept_id.

Conventions

No.	Convention Description
1	Valid Concepts for the drug_concept_id field belong to the 'Drug' domain. They are based on Ingredient Concept Class within the RxNorm Vocabulary.

For peer review only

Online supplementary material

measurement

The measurement table contains records of structured values (numerical or categorical) obtained through systematic and standardized examination or testing of a person. In the NeuroBlu dataset, the table contains measurements of:

1. Clinical Global Impression - Severity (CGI-S)
2. Clinical Global Impression - Improvement (CGI-I)
3. Global Assessment of Functioning (GAF)
4. Montgomery-Asberg Depression Rating Scale (MADRS)

Columns

Field	Type	Description
measurement_id	serial	A unique identifier for each measurement.
person_id	int8	A foreign key identifier to the person for whom the measurement is recorded. The demographic details of that person are stored in the person table.
measurement_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies. Joining on this value specified which clinical measurement is being recorded.
measurement_datetime	timestamp	The date and time of the measurement. The time is not captured in the source and defaults to 00: 00: 00.
measurement_type_concept_id	int4	A foreign key to the predefined Concept in the Standardized Vocabularies reflecting the provenance from where the Measurement record was recorded. These belong to the 'Meas Type' vocabulary.
operator_concept_id	int4	A foreign key identifier to the predefined Concept in the Standardized Vocabularies reflecting the mathematical operator that is applied to the value_as_number. Operators are <, < = , = , > = , > and these concepts belong to the 'Meas Value Operator' domain.
value_as_number	numeric	A measurement result where the result is expressed as a numeric value.
provider_id	int8	A foreign key to the primary care provider the person is seeing.
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the measurement was recorded.
measurement_source_value	varchar	Source value for measurement_concept_id.
measurement_source_concept_id	int4	A foreign key to a Concept in the Standardized Vocabularies that refers to the code used in the source.
unit_source_value	varchar	Source value for the unit.
value_source_value	varchar	Source value associated with value_as_number column.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

custom2_str	varchar	Custom field to indicate which measurement the record is for.

Conventions

No.	Convention Description
1	Measurements are stored as attribute value pairs, with the attribute as the measurement_concept_id and the value representing the result (value_as_number).
2	The operator_concept_id is optionally given for relative Measurements where the precise value is not available but its relation to a certain benchmarking value is available. For example, this can be used for minimal detection thresholds of a test.

For peer review only

Online supplementary material

observation

The observation table contains records of clinical facts about a related to medical history, the stated need for certain treatment, social circumstances, lifestyle choices, healthcare utilization patterns, etc. These records are obtained during examination or procedure and differ from Measurements in that they do not require a standardized test to generate clinical fact.

In NeuroBlu, the observation table contains the social history in the form of external environmental stressors.

Columns

Field	Type	Description
observation_period_id	serial	A unique identifier for each observation.
person_id	int8	A foreign key identifier to the person for whom the observation period is defined. The demographic details of that person are stored in the person table.
observation_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies.
observation_datetime	timestamp	The date and time of the observation. The time is not captured in the source and defaults to 00: 00: 00.
observation_type_concept_id	int4	A foreign key to the predefined Concept in the Standardized Vocabularies reflecting the provenance from where the Observation record was recorded. These belong to the 'Observation Type' vocabulary.
value_as_string	varchar(255)	Categorical value of the Result of the Observation, if applicable and available.
value_as_concept_id	int4	A foreign key to an observation result stored as a Concept ID. This is applicable to observations where the result can be expressed as a Standard Concept from the Standardized Vocabularies (e.g., positive/negative, present/absent, low/high, etc.).
qualifier_concept_id	int4	Contains all attributes specifying the clinical fact further, such as degrees, severities, drug-drug interaction alerts
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the measurement was recorded.
observation_source_value	varchar(255)	The observation code (Eg: ICD10) as it appears in the source data.
qualifier_source_value	varchar(50)	The source value associated with a qualifier to characterize the observation.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Conventions

No.	Convention Description
1	Observations can be stored as attribute value pairs, with the attribute as the Observation Concept and the value representing the clinical fact.
2	Valid Observation Concepts are not enforced to be from any domain though they still should be Standard Concepts.

External Stressors data derived using NLP

In the NeuroBlu dataset, the raw data for external environmental stressors (or social history) is a mixture of structured values and custom, user-defined free text. These values need to be normalized to structured values before they are used for downstream analytics. To achieve this, we have applied Natural Language Processing (NLP) to map the unstructured free text to structured labels.

The NLP model maps the semi-structured stessor data to one of the possible 61 labels (which can be categorized into 8 categories). Each label is also associated with a polarity (positive, negative or neutral) as described in the table below.

Category	Stressor Label	Polarity
Academic	Discord with teachers and classmates	Negative
Academic	Good achievement in school	Positive
Academic	Illiteracy and low-level literacy	Negative
Academic	Suspension/expulsion	Negative
Academic	Underachievement in school	Negative
Environment	Abusive environment at home	Negative
Environment	Chaotic atmosphere / stressful home environment	Negative
Environment	Cultural conflicts/issue	Negative
Environment	Exposure to Disaster, War, or Other Hostilities	Negative
Environment	Inadequate care (history from caregiver or patient)	Negative
Environment	Migration/ Acculturation Difficulty	Negative
Environment	Problem Related to Living Alone	Negative
Environment	Stable living situation	Positive
Family	Anxiety about sick person in family, dependent relative needing care at home	Negative
Family	Bereavement	Negative
Family	Disruption of Family by Separation or Divorce	Negative
Family	Inadequate or distorted communication within family	Negative
Family	Isolated family, lack of family support	Negative
Family	Legal problem in the family	Negative
Family	Parenting situation	Negative
Family	Problems in parent-child relationships and in-laws	Negative
Family	Sibling Relational Problem	Negative
Family	Substance abuse in the family	Negative
Family	Supportive family, love and respect in the family	Positive
Financial	Extreme Poverty/ Low income	Negative
Financial	Financial stability	Positive
Financial	Inadequate health/social Insurance or Welfare Support	Negative
Financial	Poor money management / no savings	Negative
Financial	Problems related to loan / creditors	Negative
Housing	Discord With Neighbor	Negative

For peer review only

Online supplementary material

Housing	Homelessness	Negative
Housing	Inadequate housing	Negative
Housing	Isolated dwelling (Lack of transportation)	Negative
Housing	Unstable housing arrangement	Negative
Legal	Arrest, Imprisonment and other incarceration	Negative
Legal	Child custody or support proceedings	Negative
Legal	Litigation / legal proceedings	Negative
Legal	Offense / charges	Negative
Legal	Problem with legal personnel	Negative
Legal	Restrain order	Negative
Occupational	Change of job	Neutral
Occupational	Fired	Negative
Occupational	Job dissatisfaction	Negative
Occupational	Job satisfaction	Positive
Occupational	Occupational stress	Negative
Occupational	Threat of job loss, Underperform at work	Negative
Occupational	Unemployment	Negative
Occupational	Unstable job (changing jobs often)	Negative
Personal	Emotional abuse	Negative
Personal	Involuntary treatment	Negative
Personal	Major life events	Negative
Personal	Physical abuse (spouse)	Negative
Personal	Relationship Distress With Spouse or Intimate Partner	Negative
Personal	Sense of meaning and support from belief	Positive
Personal	Sexual abuse (spouse)	Negative
Personal	Target of Adverse Discrimination	Negative
Personal	Victim of crime	Negative
Social	Conflictual social relationships	Negative
Social	Inadequate social support	Negative
Social	Lack of social circles	Negative
Social	Supportive friends	Positive

Precision statistics for the NLP classifiers provided as the median metrics of each stressor category are provided in the table below.

Category	AUROC	Recall	Precision	F1
Academic	0.999	0.667	1.000	0.800
Environment	0.999	0.923	0.800	0.857
Family	0.984	0.519	0.759	0.619
Financial	0.998	0.500	0.833	0.667
Housing	0.993	0.417	0.633	0.473
Legal	0.985	0.708	0.813	0.745
Occupational	0.993	0.500	0.895	0.664
Social	0.990	0.500	0.833	0.625

For peer review only

Online supplementary material

(ii) Family History

The family_history table contains data on recorded diagnoses of mental disorders affecting patients' relatives in the form of ICD-10-CM codes, along with the degree of relationship relative to the patient.

Columns

Field	Type	Description
family_history_id	serial	A unique identifier for each family history record.
person_id	int8	A foreign key identifier to the person for whom the MSE is recorded. The demographic details of that person are stored in the person table.
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the data was recorded.
condition_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies belonging to the 'Condition' Domain. Joining on this value extracts the corresponding ICD-9-CM or ICD-10-CM diagnostic code.
condition_source_value	varchar(255)	Source value for condition_concept_id.

For peer review only

Online supplementary material

(iii) Derived Measurements

The derived_measurement table contains records of structured values (numerical or categorical) obtained through deriving recorded measurements in the measurement table. In the NeuroBlu dataset, the table contains measurements of:

1. Smoothened CGI-S
2. Derived MADRS

Columns

Field	Type	Description
derived_measurement_id	serial	A unique identifier for each derived measurement.
person_id	int8	A foreign key identifier to the person for whom the measurement is recorded. The demographic details of that person are stored in the person table.
derived_measurement_value	float4	The value of the derived measurement.
derived_measurement_date	date	The date of the derived measurement.
range_low	numeric	The lower limit of the normal range of the measurement result. The lower range is assumed to be of the same unit of measure as the measurement value.
range_high	numeric	The upper limit of the normal range of the measurement result. The upper range is assumed to be of the same unit of measure as the measurement value.
provider_id	int8	A foreign key to the primary care provider the person is seeing.
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the measurement was recorded.
unit_source_value	varchar(50)	Source value for the unit.
custom2_str	varchar(255)	Custom field to indicate which measurement the record is for (i.e. Derived MADRS or Smoothened CGI-S)

Conventions

No.	Convention Description
1	Measurements are stored as attribute value pairs, with the attribute as the custom2_str and the value representing the result (derived_measurement_value).

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Smoothened CGI-S

In the NeuroBlu dataset, Clinical Global Impression-Severity (CGI-S) measurements are found to be noisy, both the severity and the improvement scales. The values recorded fluctuate across measurements. Sometimes a visit doesn't have any CGI-S measurement, and sometimes multiple CGI-S measurements are recorded in a single day.

To tackle this problem, we provide smoothened values of CGI-S. Raw measurement values are smoothened and interpolated on days with visit but no CGI-S measurements (the visit is recorded in visit_occurrence, but no CGI-S is measured on that day). These smoothened and interpolated values are available in the derived_measurement table.

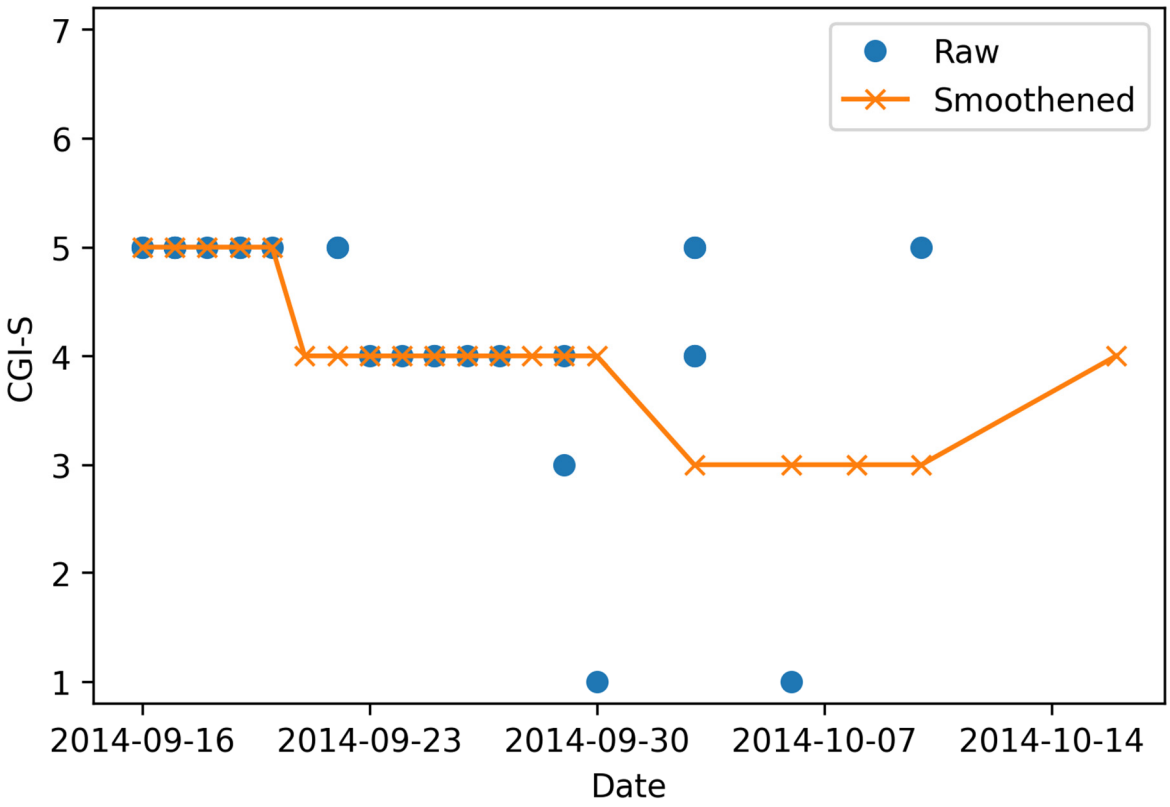
The smoothing and interpolation is done on a per-patient basis using exponential-weighted smoothing, based on the following equation:

$$s_i = \sum_{j \in J} w_{i,j} c_j$$

$$w_{i,j} = \frac{\exp(-d_{i,j}^2/(2\sigma^2))}{\sum_{j \in J} \exp(-d_{i,j}^2/(2\sigma^2))}$$

where *s_i* is the smoothened CGI-S at visit *i*, *J* is the set of all visits with CGI-S measurements, *c_j* is the measured raw CGI-S values at visit *j*, *d_i,j* is the time difference between visit *i* and *j*, in terms of days, and *σ* is the smoothing parameter, equals to 5 in our case. The equation works for days with multiple CGI measurements and days without CGI-S measurements.

The following figure shows how the smoothened CGI-S looks for an example patient.



Online supplementary material

MADRS Scores derived using CGI-S data

The Montgomery–Asberg Depression Rating Scale (MADRS) is a clinician-recorded rating scale for symptoms of depression. It comprises of 10 individual scores which are summed to yield a final score which ranges from 0 to 60. However, unlike the Clinical Global Impression-Severity (CGI-S), the NeuroBlu dataset has few documented records of MADRS scores. However, it is also one of the most widely used scores in psychiatry. We have observed that there is a fairly high correlation between the CGI-S and the final MADRS score. For this reason mapping functions have been derived that may be used to map the CGI-S score to the MADRS score.

In the event an analysis requires MADRS data, we provide values of MADRS derived from CGI-S at every timepoint with CGI-S measurement. These smoothened and interpolated values are available in the derived_measurement table.

All encounters in which both CGI-S and MADRS scores have been simultaneously measured are extracted from the dataset. In total, there are approximately 3101 such encounters. Both measurements are strongly correlated with a correlation coefficient of 0.80. This supports the possibility of mapping CGI-S to MADRS.

Upon manual inspection of the distribution of the MADRS scores for each value of CGI-S, we observed that MADRS scores have asymmetric distribution shapes for CGI-S values in the end of the spectrum, and wide distribution shapes for CGI-S values around the middle of the spectrum. For this reason, we feel that a simple linear model may not provide good confidence interval for the derived MADRS from CGI-S. We conclude that MADRS score follows the four-parameter beta distribution Beta(a, b, 0, 60), where the parameters a, b is a function of the CGI-S score, and the value is bounded between 0 and 60. Then, the values of a, b are estimated for every possible value of CGI-S using the method of moments. The derived MADRS for each CGI-S score takes value of the distribution mean, i.e. $60*a/(a+b)$. To gauge the goodness-of-fit of this approach, the R^2 value is computed and is found to be 0.66.

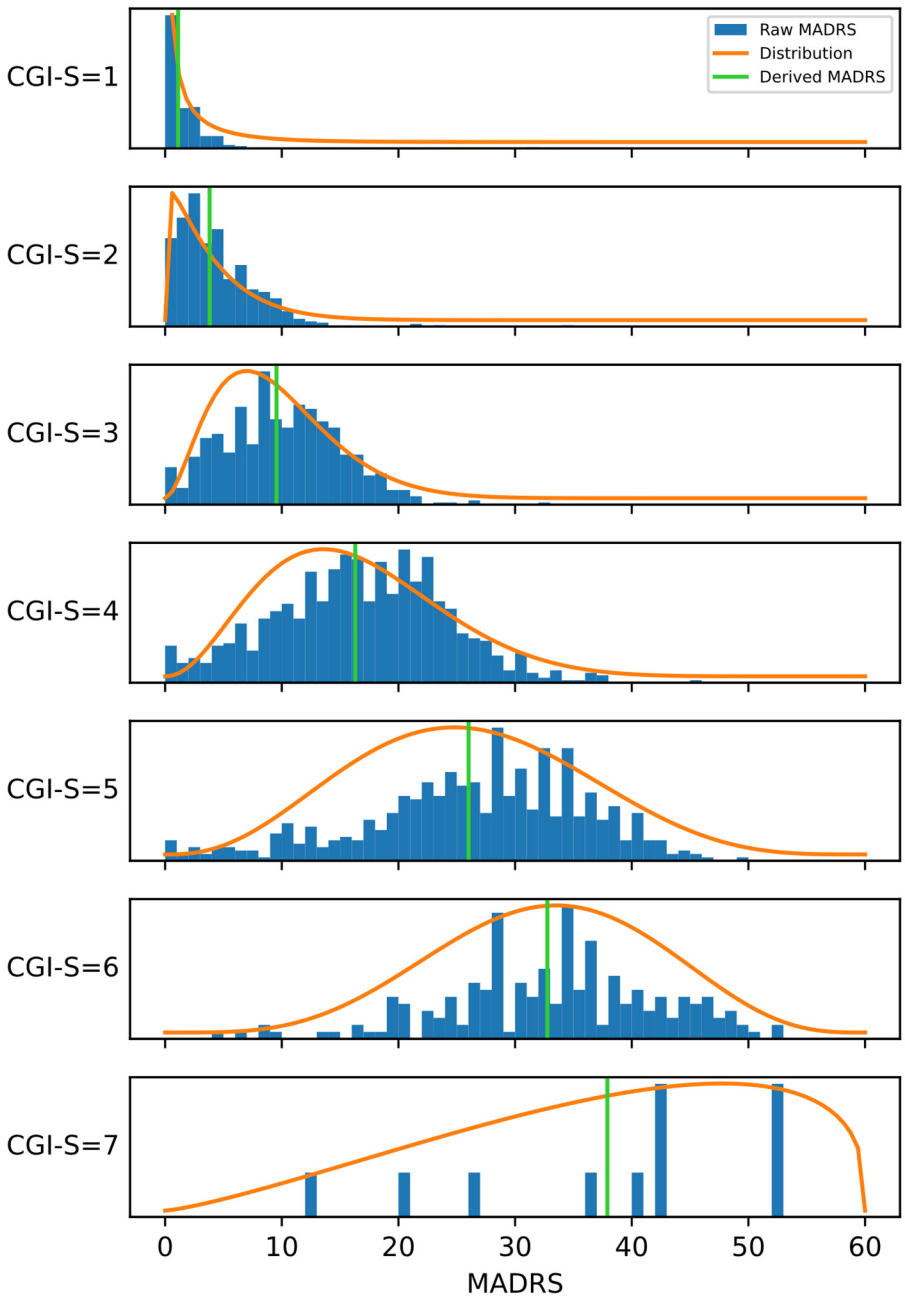
The estimated values of a and b along with the derived MADRS value for each CGI-S score are summarized in the following table.

CGI-S	a	b	Derived MADRS	95% confidence interval
1	0.1874	9.9098	1.1	0.9 to 1.3
2	1.1070	16.3081	3.8	3.6 to 4.0
3	2.8634	15.1325	9.5	9.3 to 9.8
4	3.2716	8.7725	16.3	16.0 to 16.6
5	3.7353	4.8806	26.0	25.5 to 26.6
6	5.1854	4.3100	32.8	31.9 to 33.6
7	2.2803	1.3288	37.9	32.7 to 43.1

The distributions of MADRS score for each value of CGI-S are shown in the following figures. The first figure shows the histogram of MADRS scores for each value of CGI-S and the shape of the fitted beta distributions. The second figure shows the relationship between derived MADRS and CGI-S on top of a scatter plot of raw MADRS versus raw CGI-S.

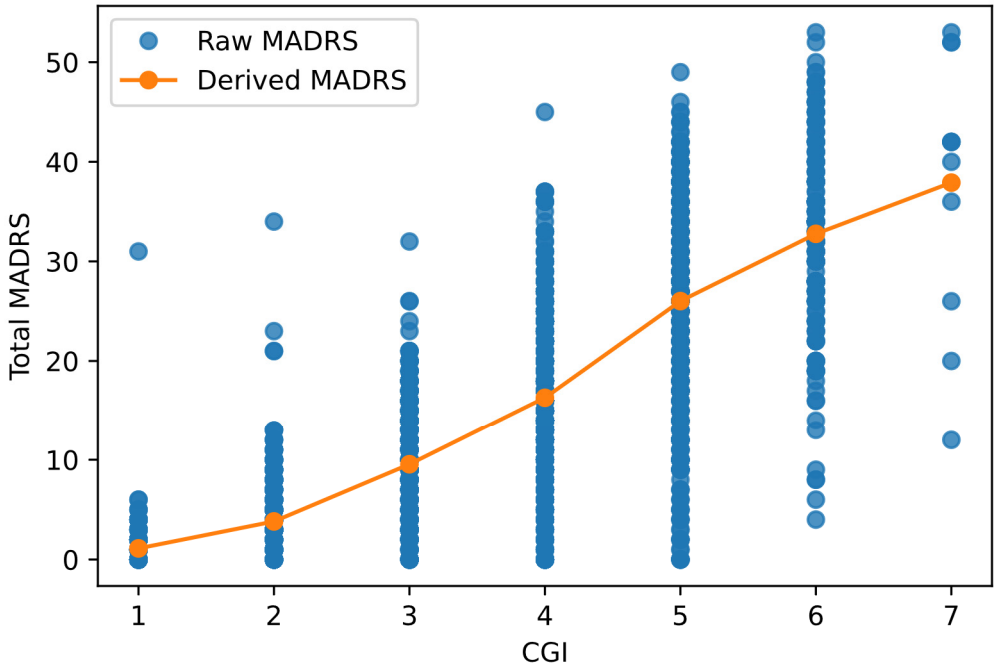
For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material



er review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

(iv) Mental State Examination

The derived_mse table contains the cleaned form of the mental state examination (MSE) records. In the NeuroBlu dataset, the MSE data are represented as Status Assessment Vectors (SAV) which are further described subsequently.

Columns

Field	Type	Description
person_id	int8	A foreign key identifier to the person for whom the MSE is recorded. The demographic details of that person are stored in the person table.
measurement_date	date	The date of the derived MSE.
Days	int4	Number of days of measurement since the patient was first recorded in the dataset.
sav_i_cat_j	int4	i-th value of the SAV. Also correspond to the j-th value for cat category. Contains binary value. i takes value from 0 to 240.

Status Assessment Vectors

In the NeuroBlu dataset, the raw data for MSE is recorded in the form of semi-structured free text, making it difficult to analyze in their raw form.

To tackle this problem, we provide natural language processing (NLP) derived structured values of MSE. Raw MSE data for a single patient on a particular day are cleaned and passed through an NLP model that returns a 241-element vector form of MSE. This vector is called the Status Assessment Vector (SAV). SAV is a binary vector, where the value is 1 if the category-factor combination of the SAV is applicable for the patient. As an example, the 15th position in the SAV corresponds to whether a patient is responding to internal stimuli or not and value of 1 for this position means that the patient is responding to internal stimuli. Further details of the NLP pipeline are described in the following article:

Mukherjee, S.S., Yu, J., Won, Y., McClay, M.J., Wang, L., Rush, A.J. and Sarkar, J., 2020. Natural Language Processing-Based Quantification of the Mental State of Psychiatric Patients. *Computational Psychiatry*, 4, pp.76–106. DOI: http://doi.org/10.1162/cpsy_a_00030

The table below maps each element of the SAV to its location in the derived_mse table.

For peer review only

Online supplementary material

SAV #	Category	Factor
0	Abnormal or psychotic thoughts	no issues
1	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (grandeur)
2	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (persecution)
3	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (religious)
4	Abnormal or psychotic thoughts	experiencing hallucinations (auditory)
5	Abnormal or psychotic thoughts	experiencing hallucinations (visual)
6	Abnormal or psychotic thoughts	experiencing hallucinations (tactile)
7	Abnormal or psychotic thoughts	History of delusions/abnormal thoughts
8	Abnormal or psychotic thoughts	History of hallucinations
9	Abnormal or psychotic thoughts	neutral/unable to categorize
10	Abnormal or psychotic thoughts	experiencing delusions (not specified)
11	Abnormal or psychotic thoughts	experiencing hallucinations (not specified)
12	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (obsessions)
13	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (paranoia)
14	Abnormal or psychotic thoughts	experiencing hallucinations (olfactory)
15	Abnormal or psychotic thoughts	responding to internal stimuli
16	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (sexual)
17	Affect	no issues/appropriate affect
18	Affect	neutral/unable to categorize
19	Affect	aggressive
20	Affect	reactive
21	Affect	irritable/angry
22	Affect	improved from last visit
23	Affect	issues with cognition
24	Affect	loss of energy/tiredness

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

25	Affect	blunted/restricted
26	Affect	anxious/fearful
27	Affect	tearful/depressed/sad
28	Affect	bright/positive
29	Affect	labile
30	Affect	intense
31	Affect	expansive/overly animated
32	Affect	generally odd or out of the ordinary/inappropriate
33	Affect	flooded
34	Appearance	Issues with Affect
35	Appearance	Sleep/Bed
36	Appearance	Issues with Eye Contact
37	Appearance	Not age appropriate
38	Appearance	Issues with general Health
39	Appearance	Issues with Behavior
40	Appearance	Positive Affect
41	Appearance	Positive Grooming/Hygiene
42	Appearance	Issues with Grooming/Hygiene
43	Appearance	Positive Health
44	Appearance	Issues with Weight
45	Appearance	Normal/no issues
46	Appearance	Hospital attire
47	Appearance	Positive dress
48	Appearance	Issues with dress
49	Appearance	Neutral
50	Appearance	Positive eye contact
51	Appearance	Appropriate Age
52	Appearance	Positive Behavior
53	Association	intact/no issues
54	Association	neutral/unable to categorize
55	Association	circumstantial
56	Association	loose
57	Association	tangential
58	Association	age appropriate
59	Association	improved
60	Association	declined
61	Association	disorganized
62	Association	general issues
63	Attention/ Concentration	intact/no issues
64	Attention/ Concentration	neutral/unable to categorize

For peer review only

Online supplementary material

65	Attention/ Concentration	general issues
66	Attention/ Concentration	improved
67	Attention/ Concentration	declined
68	Attention/ Concentration	age appropriate
69	Attention/ Concentration	issues due to MH
70	Attention/ Concentration	issues due to DD
71	Attention/ Concentration	varying
72	Attitude	positive/appropriate
73	Attitude	neutral/unable to categorize
74	Attitude	age appropriate
75	Attitude	not age appropriate
76	Attitude	negative/inappropriate
77	Attitude	improved
78	Attitude	declined
79	Cognition	normal/no issues
80	Cognition	neutral/unable to categorize
81	Cognition	issues with concentration
82	Cognition	positive concentration
83	Cognition	issues with fund of knowledge
84	Cognition	positive fund of knowledge
85	Cognition	issues with attention
86	Cognition	positive attention
87	Cognition	general issues
88	Executive Functioning	intact/no issues
89	Executive Functioning	neutral/unable to categorize
90	Executive Functioning	some impairment
91	Executive Functioning	much impairment
92	Executive Functioning	improved since last session
93	Executive Functioning	declined since last session
94	Executive Functioning	age appropriate
95	Executive Functioning	below age level
96	Executive Functioning	impaired (due to DD)

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

97	Executive Functioning	impaired (due to TBI)
98	Executive Functioning	impaired (due to other MH issue)
99	Fund of knowledge	intact/no issues
100	Fund of knowledge	neutral/unable to categorize
101	Fund of knowledge	positive vocabulary
102	Fund of knowledge	issues with vocabulary
103	Fund of knowledge	positive history
104	Fund of knowledge	issues with history
105	Fund of knowledge	average
106	Fund of knowledge	above average
107	Fund of knowledge	below average
108	Fund of knowledge	positive current events
109	Fund of knowledge	issues with current events
110	Fund of knowledge	generally limited/issues
111	Fund of knowledge	declined
112	Fund of knowledge	improved
113	Gait and station	normal/no issues
114	Gait and station	neutral/unable to categorize
115	Gait and station	non-ambulatory
116	Gait and station	issues with gait/station
117	Gait and station	assistive device needed
118	Homicidal	not present
119	Homicidal	neutral
120	Homicidal	present
121	Impulse control	good/no issues
122	Impulse control	neutral/unable to categorize
123	Impulse control	limited/some issues
124	Impulse control	poor/serious issues
125	Impulse control	age appropriate
126	Impulse control	not age appropriate
127	Insight	no issues with insight
128	Insight	issues with insight
129	Insight	neutral/unable to categorize
130	Intelligence	normal/no issues with intelligence
131	Intelligence	issues with intelligence
132	Intelligence	neutral/unable to categorize
133	Judgment	no issues with judgement
134	Judgment	issues with judgement
135	Judgment	neutral/unable to categorize
136	Language	intact
137	Language	neutral/unable to categorize

For peer review only

Online supplementary material

138	Language	repetition intact
139	Language	issues with repetition
140	Language	object naming intact
141	Language	issues with object naming
142	Language	impaired
143	Language	non-verbal/mute
144	Language	minimally verbal
145	Language	issues related to DD
146	Level of consciousness	normal; no issues notes
147	Level of consciousness	neutral; unable to categorize
148	Level of consciousness	sleepy/drowsy
149	Level of consciousness	clouded
150	Level of consciousness	issues verbally
151	Level of consciousness	misc. issues
152	Level of consciousness	impaired due to DD/MI
153	Memory	no issues/memory unimpaired
154	Memory	issues/memory impaired
155	Memory	neutral/unable to categorize
156	Mood	Normal; no issues
157	Mood	neutral; unable to categorize
158	Mood	"anxious, tense"
159	Mood	"depressed, sad, despondent"
160	Mood	"irritable, angry"
161	Mood	improved
162	Mood	declined
163	Mood	"elevated, manic"
164	Mood	labile
165	Mood	misc. issues
166	Mood	suicidal
167	Orientation	no orientation issues
168	Orientation	issues with orientation to person
169	Orientation	issues with orientation to place
170	Orientation	issues with orientation to time/date
171	Orientation	issues with orientation to situation
172	Orientation	neutral/unable to categorize
173	Orientation	general orientation issues

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

174	Psychomotor	normal; no issues
175	Psychomotor	neutral; unable to categorize
176	Psychomotor	retarded; slowed
177	Psychomotor	agitated; tense; restless
178	Psychomotor	improved
179	Psychomotor	declined
180	Psychomotor	constricted
181	Psychomotor	tremors
182	Psychomotor	tics
183	Psychomotor	involuntary movements
184	Psychomotor	misc. issues
185	Psychomotor	catatonic
186	Reasoning	normal; no issues
187	Reasoning	neutral; unable to categorize
188	Reasoning	impaired; poor
189	Reasoning	impaired due to DD/cognitive issues
190	Reasoning	impaired due to MI
191	Reasoning	concrete
192	Reasoning	abstract
193	Reasoning	improved
194	Reasoning	misc. issues
195	Sensorium	normal; no issues
196	Sensorium	neutral; unable to categorize
197	Sensorium	clouded; sedate
198	Sensorium	impaired
199	Sensorium	fluctuating
200	Sensorium	impaired due to medications/drugs/alcohol
201	Sensorium	sleepy; lethargic
202	Sleep	"normal, no issues"
203	Sleep	"neutral, unable to categorize"
204	Sleep	maintenance issues (can't stay asleep)
205	Sleep	onset issues (unable to fall asleep)
206	Sleep	nightmares
207	Sleep	insomnia
208	Sleep	hypersomnia
209	Sleep	misc. issues related to sleep
210	Speech	"normal, no issues"
211	Speech	"neutral, unable to categorize"
212	Speech	issues with rate
213	Speech	issues with tone

For peer review only

Online supplementary material

214	Speech	issues with volume
215	Speech	stutter or lisp
216	Speech	misc. issues
217	Speech	impoverished
218	Speech	improved
219	Speech	unintelligible
220	Speech	"mute, non-verbal"
221	Suicidal	"normal, no issues, not present"
222	Suicidal	"neutral, unable to categorize"
223	Suicidal	suicidal ideation
224	Suicidal	suicidal with intent
225	Suicidal	suicidal with plan
226	Suicidal	suicide attempt
227	Suicidal	history of ideation
228	Suicidal	history of attempt
229	Suicidal	self-injurious
230	Suicidal	history of self-injury
231	Suicidal	suicidal ideation with means
232	Violent Thoughts	"normal, no issues"
233	Violent Thoughts	"neutral, unable to categorize"
234	Violent Thoughts	present
235	Violent Thoughts	"present, with intent"
236	Violent Thoughts	present with plan
237	Violent Thoughts	present with means
238	Violent Thoughts	history of violent thoughts
239	Violent Thoughts	history of violent behaviors
240	Violent Thoughts	violent/aggressive behavior

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

eTable 1: Description of U.S. clinics contributing EHR data to NeuroBlu

Mental Healthcare Provider	Website	State	Years active in dataset	Patient count	Type of setting Inpatient / outpatient / residential?	Services provided (data obtained from SAMHSA directory) The SAMHSA directory has the list of all possible services an organization can provide.	Domain focus Substance Use, Mental Health / Special Programs/Groups Offered	Any specific disease focus? such as MDD, ADHD etc.
Arapahoe House	https://www.arapahoehouse.org	Colorado	2010 - 2013	45769	Outpatient	Outpatient And Intensive Outpatient 24-Hour Care Residential DUI/DUID Classes Case Management Locations & Hours Medication Assisted Treatment We Accept Health Insurance		
Aurora Mental Health Center	https://www.aumhc.org	Colorado	2006 - 2016	78321	Hospital Outpatient, Telemedicine / telehealth	Behavior modification, Cognitive behavioral therapy, Couples/family therapy, Dialectical behavior therapy, Group therapy, Integrated dual disorders treatment, Individual psychotherapy, Psychotropic medication, Trauma therapy	Seniors or older adults, Persons who have experienced trauma, Persons with post-traumatic stress disorder (PTSD), Children/adolescents with serious emotional disturbance (SED)	Persons with post-traumatic stress disorder (PTSD)
Boulder County Public Health	https://www.bouldercounty.org/departments/public-health/	Colorado	2010 - 2015	8997	Outpatient	Monitor the health status of the population and the environment, and identify community health problems, , Prevent and control the spread of communicable disease, Promote positive health behaviors and environmental practices, Mobilize community partnerships to solve identified health problems, Enforce laws and	Reduce Suicide, Improve Early Childhood Social & Emotional Development & Reduce Post-Partum Depression, Decrease Stigma & Increase Capacity	
Campagna Academy	http://www.campagnaacademy.org	Indiana	2010 - 2015	908	Residential	Activity therapy, Cognitive behavioral therapy, Couples/family therapy, Group therapy, Integrated dual disorders treatment, Individual psychotherapy, Psychotropic medication		
Center for Child and Family Health - Duke University Medical Center	https://www.ccfhnc.org/	North Carolina	2010 - 2017	7735	Inpatient	Activity therapy, Cognitive behavioral therapy, Couples/family therapy, Dialectical behavior therapy, Electroconvulsive therapy, Group therapy, Individual psychotherapy, Psychotropic medication		
CenterPointe Hospital	https://www.centerpointehospital.com/locations/	Missouri	2013 - 2016	37262	Inpatient, Partial hospitalization/day treatment, Residential, Telemedicine/telehealth	Activity therapy, Cognitive behavioral therapy, Dialectical behavior therapy, Electroconvulsive therapy, Group therapy, Integrated dual disorders treatment, Individual psychotherapy, Psychotropic medication		
Citizen Advocates	https://citizenadvocates.net	New York	2005 - 2020	7479	Hospital Outpatient	Cognitive behavioral therapy, Couples/family therapy, Dialectical behavior therapy, Group therapy, Integrated dual disorders treatment, Individual psychotherapy, Psychotropic medication, Trauma therapy		
Columbia University Medical Center Dept. of Psychiatry	https://www.columbiapsychiatry.org	New York	2012 - 2013	20495	Hospital Inpatient	Activity therapy, Behavior modification, Cognitive behavioral therapy, Dialectical behavior therapy, Group therapy, Integrated dual disorders treatment, Individual psychotherapy, Psychotropic medication, Trauma therapy		
Cornell Scott Hill	https://cornellscott.org/locations	Connecticut	1999 - 2015	377	Hospital Outpatient	Behavior modification, Cognitive behavioral therapy, Couples/family therapy, Group therapy, Integrated dual disorders treatment, Individual psychotherapy, Psychotropic medication, Trauma therapy	Persons with co-occurring mental and substance use disorders, Persons who have experienced trauma, Persons with post-traumatic stress disorder (PTSD)	Persons with post-traumatic stress disorder (PTSD)
Duke	https://www.dukehealth.org	North Carolina	2006 - 2016	77738	Hospital Inpatient	Activity therapy, Cognitive behavioral therapy, Couples/family therapy, Dialectical behavior therapy, Electroconvulsive therapy, Group therapy, Individual psychotherapy, Psychotropic medication	Transitional age young adults, Persons with co-occurring mental and substance use disorders, Persons with eating disorders	
Grant-Blackford	http://www.cornerstone.org/about.aspx	Indiana	2009 - 2013	7813	Hospital Outpatient & Residential	Activity therapy, Behavior modification, Cognitive behavioral therapy, Dialectical behavior therapy, Group therapy, Integrated dual disorders treatment, Individual psychotherapy, Psychotropic medication, Trauma therapy		
Horizon Behavioral Health	https://www.horizonbh.org	Virginia	2012 - 2016	15909	Hospital Outpatient, Partial hospitalization/day treatment, Telemedicine/telehealth	Activity therapy, Behavior modification, Cognitive behavioral therapy, Couples/family therapy, Dialectical behavior therapy, Group therapy, Integrated dual disorders treatment, Individual psychotherapy, Psychotropic medication, Trauma therapy	Transitional age young adults, Persons who have experienced trauma, Persons with post-traumatic stress disorder (PTSD), Children/adolescents with serious emotional disturbance (SED)	Persons with post-traumatic stress disorder (PTSD)
LifeSpring, Inc.	https://www.lifespringhealthsystems.org	Indiana	2010 - 2016	20361	Hospital Outpatient	Activity therapy, Behavior modification, Cognitive behavioral therapy, Couples/family therapy, Dialectical behavior therapy, Group therapy, Integrated dual disorders treatment, Individual psychotherapy, Psychotropic medication, Trauma therapy	Persons with co-occurring mental and substance use disorders, Persons with serious mental illness (SMI)	
Miami Beach Community Health Center	https://mbchc.org/	Florida	2008 - 2015	10289	Outpatient	Behavior modification, Cognitive behavioral therapy, Group therapy, Integrated dual disorders treatment, Individual psychotherapy, Psychotropic medication, Trauma therapy	Persons with co-occurring mental and substance use disorders, Persons with serious mental illness (SMI)	
Montefiore	https://www.montefiore.org/	New York	2012 - 2014	2100	Hospital Outpatient	Activity therapy, Behavior modification, Cognitive behavioral therapy, Couples/family therapy, Dialectical behavior therapy, Group therapy, Integrated dual disorders treatment, Individual psychotherapy, Psychotropic medication, Trauma therapy	Persons with co-occurring mental and substance use disorders, Persons who have experienced trauma, Persons with post-traumatic stress disorder (PTSD), Persons with serious mental illness (SMI)	Persons with post-traumatic stress disorder (PTSD)
Montefiore Moses and Wakefield	https://www.montefiore.org/psychiatry-contact	New York	2006 - 2016	31671	Outpatient	Activity therapy, Behavior modification, Cognitive behavioral therapy, Couples/family therapy, Dialectical behavior therapy, Group therapy, Integrated dual disorders treatment, Individual psychotherapy, Psychotropic medication, Trauma therapy	Persons with co-occurring mental and substance use disorders, Persons who have experienced trauma, Persons with post-traumatic stress disorder (PTSD), Persons with serious mental illness (SMI)	Persons with post-traumatic stress disorder (PTSD)
New River Valley Community Services	http://www.nrvcs.org	Virginia	2013 - 2013	5845	Hospital Outpatient, Telemedicine / telehealth	Behavior modification, Cognitive behavioral therapy, Couples/family therapy, Group therapy, Integrated dual disorders treatment, Individual psychotherapy, Psychotropic medication, Trauma therapy	Persons with co-occurring mental and substance use disorders, Persons who have experienced trauma, Persons with post-traumatic stress disorder (PTSD), Children/adolescents with serious emotional disturbance (SED), Persons with serious mental illness (SMI)	Persons with post-traumatic stress disorder (PTSD)
Porter-Starke Services Inc	https://www.porterstarke.org	Indiana	2005 - 2017	45280	Hospital Outpatient, Telemedicine / telehealth	Activity therapy, Behavior modification, Cognitive behavioral therapy, Dialectical behavior therapy, Group therapy, Integrated dual disorders treatment, Individual psychotherapy, Psychotropic medication, Trauma therapy	Children/adolescents with serious emotional disturbance (SED), Persons with serious mental illness (SMI)	
Psych Care Consultants (PCC)	https://www.bhsmo.com/contact-us	Missouri	2012 - 2020	63427	Outpatient	Full spectrum of comprehensive mental health services for individuals of all ages including billing services, injection clinic, ketamine infusion, medication management, neurology, psychotherapy/counseling, psych testing, pharmacy services, research, andTMS therapy		difficult to treat cases of depression, anxiety, and bipolar disorders, among others. diagnoses of Major Depressive Disorder (MDD) and Obsessive
Rimrock Foundation	https://www.rimrock.org	Montana	2013 - 2016	5164	Outpatient, Telehealth	Inpatient Rehab, Day Treatment, Outpatient Services, Family Therapy, Chemical Dependency Assessments, DUI School, Medical Detoxification, Long Term Residential for Mothers with Children, Re-Entry Programs, Spanish Speaking counselor	In addition to mental health and substance use counseling, Mental Health Services also includes, Substance Abuse Professional (SAP) services, Mental Health assessment services, Clinical supervision for licensure., EMDR, Family therapy, Grief therapy, Therapeutic	Substance Use Disorders, Sexual Addiction, Co-Dependency, Co-Occurring Disorders. Persons with post-traumatic stress disorder (PTSD)
The Bridge, Inc	http://bridgeinc.org	Alabama	2013 - 2016	3387	Outpatient & Residential Treatment	Outpatient, Residential Treatment & community based		substance use disorder treatment and behavioral rehabilitation for adolescents
Thresholds, Inc	https://yourfirststep.org/treatment-center/thresholds-inc-georgetown-de/	Delaware	2011 - 2017	7710	Outpatient	Drug & Alcohol Addiction Treatment, Drug & Alcohol Detoxification Program, Offers Buprenorphine detox, Relapse prevention with the use of naltrxone, Offers Buprenorphine used in treatment, Oral form of Naltrexone, Provides Injectable Naltrexone (Vivitrol), Clients accepted on opioid/methadone medication,		
University of Colorado - UC Health	https://www.uchealth.org/about/community-benefit-report/behavioral-health/	Colorado	2007 - 2014	2204	Hospital Inpatient, Hospital Outpatient, Telemedicine / telehealth	Counseling, Suicide prevention, Outreach	Learning Disabilities, Attention Deficit Hyperactivity Disorder (ADHD) and Attention Deficit Disorder (ADD), Behavior Disorders, Academic Achievement, Psychiatric Disorders, Intellectual Disabilities, Gifted and Talented, Autism	
West Central Behavioral Health	http://www.wcbh.org/contact/	New Hampshire	2009 - 2015	10041	Hospital Outpatient, Telemedicine / telehealth	Cognitive behavioral therapy, Couples/family therapy, Dialectical behavior therapy, Integrated dual disorders treatment, Individual psychotherapy, Psychotropic medication, Trauma therapy	Seniors or older adults, Active duty military, Military families, Persons with co-occurring mental and substance use disorders, Persons who have experienced trauma, Persons with post-traumatic stress disorder (PTSD), Children/adolescents with serious emotional disturbance	Persons with post-traumatic stress disorder (PTSD)
Westchester Medical Center	https://www.westchestermedicalcenter.org	New York	2005 - 2020	43066	Hospital Inpatient, Hospital Outpatient, Telemedicine / telehealth	Activity therapy, Cognitive behavioral therapy, Couples/family therapy, Dialectical behavior therapy, Group therapy, Individual psychotherapy, Psychotropic medication, Trauma therapy	Transitional age young adults, Seniors or older adults, Lesbian, gay, bisexual, or transgender (LGBT) clients, Persons who have experienced trauma, Persons with Alzheimer's or dementia, Children/adolescents with serious emotional disturbance (SED),	

For peer review only

Online supplementary material

eTable 2: NeuroBlu dataset race distribution

Race	Number of patients	Percentage
White	276,633	49.1%
Unknown	186,517	33.1%
Black or African American	73,504	13.1%
Other Race	13,179	2.3%
Native Hawaiian or Other Pacific Islander	6,712	1.2%
Asian	3,958	0.7%
American Indian or Alaska Native	2,437	0.4%

eTable 3: NeuroBlu dataset age distribution

Age (years)	Number of patients	Percentage
0-9	45,050	8.0%
10-19	116,741	20.7%
20-29	108,863	19.3%
30-39	94,114	16.7%
40-49	81,556	14.5%
50-59	65,498	11.6%
60-69	30,945	5.5%
70-79	15,339	2.7%
80+	4,828	0.9%

eTable 4: NeuroBlu dataset diagnosis distribution

Diagnosis	Number of patients	Percentage
Dementia & Alzheimer's disease	5,029	0.9%
Substance-use disorders	152,790	28.1%
Schizophrenia	14,592	2.7%
Schizoaffective disorder	15,044	2.8%
Bipolar disorder	69,607	12.8%
Major depressive disorder (MDD)	129,120	23.7%
Anxiety disorders	103,923	19.1%
Attention-deficit hyperactivity disorder (ADHD)	53,744	9.9%

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

eTable 5: Gender distribution by diagnosis

Diagnosis	Number of male patients	Number of female patients	Number with unknown gender	Percentage Male	Percentage Female	Percentage Unknown	Total number
Dementia & Alzheimer's disease	2,035	2,962	32	40.5%	58.9%	0.8%	5,029
Substance-use disorders	96,180	56,328	282	62.9%	36.9%	0.0%	152,790
Schizophrenia	9,163	5,401	28	62.8%	37.0%	0.4%	14,592
Schizoaffective disorder	7,807	7,222	15	51.9%	48.0%	0.3%	15,044
Bipolar disorder	27,877	41,642	88	40.0%	59.8%	0.1%	69,607
Major depressive disorder (MDD)	46,603	82,395	122	36.1%	63.8%	0.0%	129,120
Anxiety disorders	38,255	65,569	99	36.8%	63.1%	0.0%	103,923
Attention-deficit hyperactivity disorder (ADHD)	32,781	20,933	30	61.0%	38.9%	0.1%	53,744

For peer review only

Online supplementary material

eTable 6: ICD-9/ICD-10 code definitions for diagnostic category

Diagnosis	ICD code	ICD type	ICD Description
ADHD	314	ICD-9	Attention deficit disorder without mention of hyperactivity
ADHD	314.01	ICD-9	Attention deficit disorder with hyperactivity
ADHD	F90.0	ICD-10	Attention-deficit hyperactivity disorder, predominantly inattentive type
ADHD	F90.1	ICD-10	Attention-deficit hyperactivity disorder, predominantly hyperactive type
ADHD	F90.2	ICD-10	Attention-deficit hyperactivity disorder, combined type
ADHD	F90.8	ICD-10	Attention-deficit hyperactivity disorder, other type
ADHD	F90.9	ICD-10	Attention-deficit hyperactivity disorder, unspecified type
Anxiety disorders	300.01	ICD-9	Panic disorder without agoraphobia
Anxiety disorders	300.02	ICD-9	Generalized anxiety disorder
Anxiety disorders	300.2	ICD-9	Phobia, unspecified
Anxiety disorders	300.21	ICD-9	Agoraphobia with panic disorder
Anxiety disorders	300.22	ICD-9	Agoraphobia without mention of panic attacks
Anxiety disorders	300.23	ICD-9	Social phobia
Anxiety disorders	300.29	ICD-9	Other isolated or specific phobias
Anxiety disorders	300.3	ICD-9	Obsessive-compulsive disorders
Anxiety disorders	309.81	ICD-9	Posttraumatic stress disorder
Anxiety disorders	F40.00	ICD-10	Agoraphobia, unspecified
Anxiety disorders	F40.01	ICD-10	Agoraphobia with panic disorder
Anxiety disorders	F40.02	ICD-10	Agoraphobia without panic disorder
Anxiety disorders	F40.10	ICD-10	Social phobia, unspecified
Anxiety disorders	F40.11	ICD-10	Social phobia, generalized
Anxiety disorders	F40.210	ICD-10	Arachnophobia
Anxiety disorders	F40.218	ICD-10	Other animal type phobia
Anxiety disorders	F40.220	ICD-10	Fear of thunderstorms
Anxiety disorders	F40.228	ICD-10	Other natural environment type phobia

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Anxiety disorders	F40.230	ICD-10	Fear of blood
Anxiety disorders	F40.231	ICD-10	Fear of injections and transfusions
Anxiety disorders	F40.232	ICD-10	Fear of other medical care
Anxiety disorders	F40.233	ICD-10	Fear of injury
Anxiety disorders	F40.240	ICD-10	Claustrophobia
Anxiety disorders	F40.241	ICD-10	Acrophobia
Anxiety disorders	F40.242	ICD-10	Fear of bridges
Anxiety disorders	F40.243	ICD-10	Fear of flying
Anxiety disorders	F40.248	ICD-10	Other situational type phobia
Anxiety disorders	F40.290	ICD-10	Androphobia
Anxiety disorders	F40.291	ICD-10	Gynephobia
Anxiety disorders	F40.298	ICD-10	Other specified phobia
Anxiety disorders	F40.8	ICD-10	Other phobic anxiety disorders
Anxiety disorders	F40.9	ICD-10	Phobic anxiety disorder, unspecified
Anxiety disorders	F41.0	ICD-10	Panic disorder [episodic paroxysmal anxiety]
Anxiety disorders	F41.1	ICD-10	Generalized anxiety disorder
Anxiety disorders	F42.2	ICD-10	Mixed obsessional thoughts and acts
Anxiety disorders	F42.3	ICD-10	Hoarding disorder
Anxiety disorders	F42.4	ICD-10	Excoriation (skin-picking) disorder
Anxiety disorders	F42.8	ICD-10	Other obsessive-compulsive disorder
Anxiety disorders	F42.9	ICD-10	Obsessive-compulsive disorder, unspecified
Anxiety disorders	F43.10	ICD-10	Post-traumatic stress disorder, unspecified
Anxiety disorders	F43.11	ICD-10	Post-traumatic stress disorder, acute
Anxiety disorders	F43.12	ICD-10	Post-traumatic stress disorder, chronic
Bipolar Disorder	296	ICD-9	Bipolar I disorder, single manic episode, unspecified

For peer review only

Online supplementary material

Bipolar Disorder	296.01	ICD-10	Bipolar I disorder, single manic episode, mild
Bipolar Disorder	296.02	ICD-11	Bipolar I disorder, single manic episode, moderate
Bipolar Disorder	296.03	ICD-12	Bipolar I disorder, single manic episode, severe, without mention of psychotic behavior
Bipolar Disorder	296.04	ICD-13	Bipolar I disorder, single manic episode, specified as with psychotic behavior
Bipolar Disorder	296.05	ICD-14	Bipolar I disorder, single manic episode, in partial or unspecified remission
Bipolar Disorder	296.06	ICD-15	Bipolar I disorder, single manic episode, in full remission
Bipolar Disorder	296.4	ICD-9	Bipolar I disorder, most recent episode (or current) manic, unspecified
Bipolar Disorder	296.41	ICD-9	Bipolar I disorder, most recent episode (or current) manic, mild
Bipolar Disorder	296.42	ICD-9	Bipolar I disorder, most recent episode (or current) manic, moderate
Bipolar Disorder	296.43	ICD-9	Bipolar I disorder, most recent episode (or current) manic, severe, without mention of psychotic behavior
Bipolar Disorder	296.44	ICD-9	Bipolar I disorder, most recent episode (or current) manic, severe, specified as with psychotic behavior
Bipolar Disorder	296.45	ICD-9	Bipolar I disorder, most recent episode (or current) manic, in partial or unspecified remission
Bipolar Disorder	296.46	ICD-9	Bipolar I disorder, most recent episode (or current) manic, in full remission
Bipolar Disorder	296.5	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, unspecified
Bipolar Disorder	296.51	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, mild
Bipolar Disorder	296.52	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, moderate
Bipolar Disorder	296.53	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, severe, without mention of psychotic behavior
Bipolar Disorder	296.54	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, severe, specified as with psychotic behavior
Bipolar Disorder	296.55	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, in partial or unspecified remission
Bipolar Disorder	296.56	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, in full remission
Bipolar Disorder	296.6	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, unspecified
Bipolar Disorder	296.61	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, mild
Bipolar Disorder	296.62	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, moderate
Bipolar Disorder	296.63	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, severe, without mention of psychotic behavior
Bipolar Disorder	296.64	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, severe, specified as with psychotic behavior

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Bipolar Disorder	296.65	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, in partial or unspecified remission
Bipolar Disorder	296.66	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, in full remission
Bipolar Disorder	296.7	ICD-9	Bipolar I disorder, most recent episode (or current) unspecified
Bipolar Disorder	296.8	ICD-9	Bipolar disorder, unspecified
Bipolar Disorder	296.89	ICD-9	Other bipolar disorders
Bipolar Disorder	F31.0	ICD-10	Bipolar disorder, current episode hypomanic
Bipolar Disorder	F31.10	ICD-10	Bipolar disorder, current episode manic without psychotic features, unspecified
Bipolar Disorder	F31.11	ICD-10	Bipolar disorder, current episode manic without psychotic features, mild
Bipolar Disorder	F31.12	ICD-10	Bipolar disorder, current episode manic without psychotic features, moderate
Bipolar Disorder	F31.13	ICD-10	Bipolar disorder, current episode manic without psychotic features, severe
Bipolar Disorder	F31.2	ICD-10	Bipolar disorder, current episode manic severe with psychotic features
Bipolar Disorder	F31.30	ICD-10	Bipolar disorder, current episode depressed, mild or moderate severity, unspecified
Bipolar Disorder	F31.31	ICD-10	Bipolar disorder, current episode depressed, mild
Bipolar Disorder	F31.32	ICD-10	Bipolar disorder, current episode depressed, moderate
Bipolar Disorder	F31.4	ICD-10	Bipolar disorder, current episode depressed, severe, without psychotic features
Bipolar Disorder	F31.5	ICD-10	Bipolar disorder, current episode depressed, severe, with psychotic features
Bipolar Disorder	F31.60	ICD-10	Bipolar disorder, current episode mixed, unspecified
Bipolar Disorder	F31.61	ICD-10	Bipolar disorder, current episode mixed, mild
Bipolar Disorder	F31.62	ICD-10	Bipolar disorder, current episode mixed, moderate
Bipolar Disorder	F31.63	ICD-10	Bipolar disorder, current episode mixed, severe, without psychotic features
Bipolar Disorder	F31.64	ICD-10	Bipolar disorder, current episode mixed, severe, with psychotic features
Bipolar Disorder	F31.70	ICD-10	Bipolar disorder, currently in remission, most recent episode unspecified
Bipolar Disorder	F31.71	ICD-10	Bipolar disorder, in partial remission, most recent episode hypomanic
Bipolar Disorder	F31.72	ICD-10	Bipolar disorder, in full remission, most recent episode hypomanic
Bipolar Disorder	F31.73	ICD-10	Bipolar disorder, in partial remission, most recent episode manic

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Bipolar Disorder	F31.74	ICD-10	Bipolar disorder, in full remission, most recent episode manic
Bipolar Disorder	F31.75	ICD-10	Bipolar disorder, in partial remission, most recent episode depressed
Bipolar Disorder	F31.76	ICD-10	Bipolar disorder, in full remission, most recent episode depressed
Bipolar Disorder	F31.77	ICD-10	Bipolar disorder, in partial remission, most recent episode mixed
Bipolar Disorder	F31.78	ICD-10	Bipolar disorder, in full remission, most recent episode mixed
Bipolar Disorder	F31.81	ICD-10	Bipolar II disorder
Bipolar Disorder	F31.89	ICD-10	Other bipolar disorder
Bipolar Disorder	F31.9	ICD-10	Bipolar disorder, unspecified
Dementia and Alzheimer's Disease	290	ICD-9	Senile dementia, uncomplicated
Dementia and Alzheimer's Disease	290.1	ICD-9	Presenile dementia
Dementia and Alzheimer's Disease	290.2	ICD-9	Senile dementia with delusional or depressive features
Dementia and Alzheimer's Disease	290.3	ICD-9	Senile dementia with delirium
Dementia and Alzheimer's Disease	290.4	ICD-9	Vascular dementia
Dementia and Alzheimer's Disease	290.43	ICD-9	Vascular dementia, with depressed mood
Dementia and Alzheimer's Disease	294.1	ICD-9	Dementia in conditions classified elsewhere
Dementia and Alzheimer's Disease	294.8	ICD-9	Other persistent mental disorders due to conditions classified elsewhere
Dementia and Alzheimer's Disease	331	ICD-9	Alzheimer's disease
Dementia and Alzheimer's Disease	F03.90	ICD-10	Unspecified dementia without behavioral disturbance
Dementia and Alzheimer's Disease	F03.91	ICD-10	Unspecified dementia with behavioral disturbance

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Dementia and Alzheimer's Disease	G30.1	ICD-10	Alzheimer's disease with late onset
Dementia and Alzheimer's Disease	G30.8	ICD-10	Other Alzheimer's disease
MDD	296.2	ICD-9	Major depressive affective disorder, single episode, unspecified
MDD	296.21	ICD-9	Major depressive affective disorder, single episode, mild
MDD	296.22	ICD-9	Major depressive affective disorder, single episode, moderate
MDD	296.23	ICD-9	Major depressive affective disorder, single episode, severe, without mention of psychotic behavior
MDD	296.24	ICD-9	Major depressive affective disorder, single episode, severe, specified as with psychotic behavior
MDD	296.25	ICD-9	Major depressive affective disorder, single episode, in partial or unspecified remission
MDD	296.26	ICD-9	Major depressive affective disorder, single episode, in full remission
MDD	296.3	ICD-9	Major depressive affective disorder, recurrent episode, unspecified
MDD	296.31	ICD-9	Major depressive affective disorder, recurrent episode, mild
MDD	296.32	ICD-9	Major depressive affective disorder, recurrent episode, moderate
MDD	296.33	ICD-9	Major depressive affective disorder, recurrent episode, severe, without mention of psychotic behavior
MDD	296.34	ICD-9	Major depressive affective disorder, recurrent episode, severe, specified as with psychotic behavior
MDD	296.35	ICD-9	Major depressive affective disorder, recurrent episode, in partial or unspecified remission
MDD	296.36	ICD-9	Major depressive affective disorder, recurrent episode, in full remission
MDD	F32.0	ICD-10	Major depressive disorder, single episode, mild
MDD	F32.1	ICD-10	Major depressive disorder, single episode, moderate
MDD	F32.2	ICD-10	Major depressive disorder, single episode, severe without psychotic features
MDD	F32.3	ICD-10	Major depressive disorder, single episode, severe with psychotic features
MDD	F32.4	ICD-10	Major depressive disorder, single episode, in partial remission
MDD	F32.5	ICD-10	Major depressive disorder, single episode, in full remission
MDD	F32.81	ICD-10	Premenstrual dysphoric disorder
MDD	F32.89	ICD-10	Other specified depressive episodes
MDD	F32.9	ICD-10	Major depressive disorder, single episode, unspecified
MDD	F33.0	ICD-10	Major depressive disorder, recurrent, mild
MDD	F33.1	ICD-10	Major depressive disorder, recurrent, moderate

For peer review only

Online supplementary material

MDD	F33.2	ICD-10	Major depressive disorder, recurrent severe without psychotic features
MDD	F33.3	ICD-10	Major depressive disorder, recurrent, severe with psychotic symptoms
MDD	F33.40	ICD-10	Major depressive disorder, recurrent, in remission, unspecified
MDD	F33.41	ICD-10	Major depressive disorder, recurrent, in partial remission
MDD	F33.42	ICD-10	Major depressive disorder, recurrent, in full remission
MDD	F33.8	ICD-10	Other recurrent depressive disorders
MDD	F33.9	ICD-10	Major depressive disorder, recurrent, unspecified
Schizoaffective disorder	295.7	ICD-9	Schizoaffective disorder, unspecified
Schizoaffective disorder	295.71	ICD-9	Schizoaffective disorder, subchronic
Schizoaffective disorder	295.72	ICD-9	Schizoaffective disorder, chronic
Schizoaffective disorder	295.73	ICD-9	Schizoaffective disorder, subchronic with acute exacerbation
Schizoaffective disorder	295.74	ICD-9	Schizoaffective disorder, chronic with acute exacerbation
Schizoaffective disorder	295.75	ICD-9	Schizoaffective disorder, in remission
Schizoaffective disorder	F25.0	ICD-10	Schizoaffective disorder, bipolar type
Schizoaffective disorder	F25.1	ICD-10	Schizoaffective disorder, depressive type
Schizoaffective disorder	F25.8	ICD-10	Other schizoaffective disorders
Schizoaffective disorder	F25.9	ICD-10	Schizoaffective disorder, unspecified
Schizophrenia	295	ICD-9	Simple type schizophrenia, unspecified
Schizophrenia	295.01	ICD-9	Simple type schizophrenia, subchronic
Schizophrenia	295.02	ICD-9	Simple type schizophrenia, chronic
Schizophrenia	295.03	ICD-9	Simple type schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.04	ICD-9	Simple type schizophrenia, chronic with acute exacerbation
Schizophrenia	295.05	ICD-9	Simple type schizophrenia, in remission
Schizophrenia	295.1	ICD-9	Disorganized type schizophrenia, unspecified
Schizophrenia	295.11	ICD-9	Disorganized type schizophrenia, subchronic
Schizophrenia	295.12	ICD-9	Disorganized type schizophrenia, chronic
Schizophrenia	295.13	ICD-9	Disorganized type schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.14	ICD-9	Disorganized type schizophrenia, chronic with acute exacerbation
Schizophrenia	295.15	ICD-9	Disorganized type schizophrenia, in remission
Schizophrenia	295.2	ICD-9	Catatonic type schizophrenia, unspecified

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Schizophrenia	295.21	ICD-9	Catatonic type schizophrenia, subchronic
Schizophrenia	295.22	ICD-9	Catatonic type schizophrenia, chronic
Schizophrenia	295.23	ICD-9	Catatonic type schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.24	ICD-9	Catatonic type schizophrenia, chronic with acute exacerbation
Schizophrenia	295.25	ICD-9	Catatonic type schizophrenia, in remission
Schizophrenia	295.3	ICD-9	Paranoid type schizophrenia, unspecified
Schizophrenia	295.31	ICD-9	Paranoid type schizophrenia, subchronic
Schizophrenia	295.32	ICD-9	Paranoid type schizophrenia, chronic
Schizophrenia	295.33	ICD-9	Paranoid type schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.34	ICD-9	Paranoid type schizophrenia, chronic with acute exacerbation
Schizophrenia	295.35	ICD-9	Paranoid type schizophrenia, in remission
Schizophrenia	295.4	ICD-9	Schizophreniform disorder, unspecified
Schizophrenia	295.41	ICD-9	Schizophreniform disorder, subchronic
Schizophrenia	295.42	ICD-9	Schizophreniform disorder, chronic
Schizophrenia	295.43	ICD-9	Schizophreniform disorder, subchronic with acute exacerbation
Schizophrenia	295.44	ICD-9	Schizophreniform disorder, chronic with acute exacerbation
Schizophrenia	295.45	ICD-9	Schizophreniform disorder, in remission
Schizophrenia	295.5	ICD-9	Latent schizophrenia, unspecified
Schizophrenia	295.51	ICD-9	Latent schizophrenia, subchronic
Schizophrenia	295.52	ICD-9	Latent schizophrenia, chronic
Schizophrenia	295.53	ICD-9	Latent schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.54	ICD-9	Latent schizophrenia, chronic with acute exacerbation
Schizophrenia	295.55	ICD-9	Latent schizophrenia, in remission
Schizophrenia	295.6	ICD-9	Schizophrenic disorders, residual type, unspecified
Schizophrenia	295.61	ICD-9	Schizophrenic disorders, residual type, subchronic
Schizophrenia	295.62	ICD-9	Schizophrenic disorders, residual type, chronic
Schizophrenia	295.63	ICD-9	Schizophrenic disorders, residual type, subchronic with acute exacerbation
Schizophrenia	295.64	ICD-9	Schizophrenic disorders, residual type, chronic with acute exacerbation
Schizophrenia	295.65	ICD-9	Schizophrenic disorders, residual type, in remission
Schizophrenia	295.8	ICD-9	Other specified types of schizophrenia, unspecified
Schizophrenia	295.81	ICD-9	Other specified types of schizophrenia, subchronic
Schizophrenia	295.82	ICD-9	Other specified types of schizophrenia, chronic
Schizophrenia	295.83	ICD-9	Other specified types of schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.84	ICD-9	Other specified types of schizophrenia, chronic with acute exacerbation
Schizophrenia	295.85	ICD-9	Other specified types of schizophrenia, in remission

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Schizophrenia	295.9	ICD-9	Unspecified schizophrenia, unspecified
Schizophrenia	295.91	ICD-9	Unspecified schizophrenia, subchronic
Schizophrenia	295.92	ICD-9	Unspecified schizophrenia, chronic
Schizophrenia	295.93	ICD-9	Unspecified schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.94	ICD-9	Unspecified schizophrenia, chronic with acute exacerbation
Schizophrenia	295.95	ICD-9	Unspecified schizophrenia, in remission
Schizophrenia	F20.0	ICD-10	Paranoid schizophrenia
Schizophrenia	F20.1	ICD-10	Disorganized schizophrenia
Schizophrenia	F20.2	ICD-10	Catatonic schizophrenia
Schizophrenia	F20.3	ICD-10	Undifferentiated schizophrenia
Schizophrenia	F20.5	ICD-10	Residual schizophrenia
Schizophrenia	F20.81	ICD-10	Schizophreniform disorder
Schizophrenia	F20.89	ICD-10	Other schizophrenia
Schizophrenia	F20.9	ICD-10	Schizophrenia, unspecified
Substance-use -disorders	291	ICD-9	Alcohol withdrawal delirium
Substance-use -disorders	291.1	ICD-9	Alcohol-induced persisting amnestic disorder
Substance-use -disorders	291.2	ICD-9	Alcohol-induced persisting dementia
Substance-use -disorders	291.3	ICD-9	Alcohol-induced psychotic disorder with hallucinations
Substance-use -disorders	291.4	ICD-9	Idiosyncratic alcohol intoxication
Substance-use -disorders	291.5	ICD-9	Alcohol-induced psychotic disorder with delusions
Substance-use -disorders	291.81	ICD-9	Alcohol withdrawal
Substance-use -disorders	291.82	ICD-9	Alcohol induced sleep disorders
Substance-use -disorders	291.89	ICD-9	Other alcohol-induced mental disorders
Substance-use -disorders	291.9	ICD-9	Unspecified alcohol-induced mental disorders
Substance-use -disorders	292	ICD-9	Drug withdrawal
Substance-use -disorders	292.11	ICD-9	Drug-induced psychotic disorder with delusions
Substance-use -disorders	292.12	ICD-9	Drug-induced psychotic disorder with hallucinations
Substance-use -disorders	292.2	ICD-9	Pathological drug intoxication
Substance-use -disorders	292.81	ICD-9	Drug-induced delirium
Substance-use -disorders	292.82	ICD-9	Drug-induced persisting dementia

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	292.83	ICD-9	Drug-induced persisting amnestic disorder
Substance-use -disorders	292.84	ICD-9	Drug-induced mood disorder
Substance-use -disorders	292.85	ICD-9	Drug induced sleep disorders
Substance-use -disorders	292.89	ICD-9	Other specified drug-induced mental disorders
Substance-use -disorders	292.9	ICD-9	Unspecified drug-induced mental disorder
Substance-use -disorders	303	ICD-9	Acute alcoholic intoxication in alcoholism, unspecified
Substance-use -disorders	303.01	ICD-9	Acute alcoholic intoxication in alcoholism, continuous
Substance-use -disorders	303.02	ICD-9	Acute alcoholic intoxication in alcoholism, episodic
Substance-use -disorders	303.03	ICD-9	Acute alcoholic intoxication in alcoholism, in remission
Substance-use -disorders	303.9	ICD-9	Other and unspecified alcohol dependence, unspecified
Substance-use -disorders	303.91	ICD-9	Other and unspecified alcohol dependence, continuous
Substance-use -disorders	303.92	ICD-9	Other and unspecified alcohol dependence, episodic
Substance-use -disorders	303.93	ICD-9	Other and unspecified alcohol dependence, in remission
Substance-use -disorders	304	ICD-9	Opioid type dependence, unspecified
Substance-use -disorders	304.01	ICD-9	Opioid type dependence, continuous
Substance-use -disorders	304.02	ICD-9	Opioid type dependence, episodic
Substance-use -disorders	304.03	ICD-9	Opioid type dependence, in remission
Substance-use -disorders	304.1	ICD-9	Sedative, hypnotic or anxiolytic dependence, unspecified
Substance-use -disorders	304.11	ICD-9	Sedative, hypnotic or anxiolytic dependence, continuous
Substance-use -disorders	304.12	ICD-9	Sedative, hypnotic or anxiolytic dependence, episodic
Substance-use -disorders	304.13	ICD-9	Sedative, hypnotic or anxiolytic dependence, in remission
Substance-use -disorders	304.2	ICD-9	Cocaine dependence, unspecified
Substance-use -disorders	304.21	ICD-9	Cocaine dependence, continuous
Substance-use -disorders	304.22	ICD-9	Cocaine dependence, episodic
Substance-use -disorders	304.23	ICD-9	Cocaine dependence, in remission

For peer review only

Online supplementary material

Substance-use -disorders	304.3	ICD-9	Cannabis dependence, unspecified
Substance-use -disorders	304.31	ICD-9	Cannabis dependence, continuous
Substance-use -disorders	304.32	ICD-9	Cannabis dependence, episodic
Substance-use -disorders	304.33	ICD-9	Cannabis dependence, in remission
Substance-use -disorders	304.4	ICD-9	Amphetamine and other psychostimulant dependence, unspecified
Substance-use -disorders	304.41	ICD-9	Amphetamine and other psychostimulant dependence, continuous
Substance-use -disorders	304.42	ICD-9	Amphetamine and other psychostimulant dependence, episodic
Substance-use -disorders	304.43	ICD-9	Amphetamine and other psychostimulant dependence, in remission
Substance-use -disorders	304.5	ICD-9	Hallucinogen dependence, unspecified
Substance-use -disorders	304.51	ICD-9	Hallucinogen dependence, continuous
Substance-use -disorders	304.52	ICD-9	Hallucinogen dependence, episodic
Substance-use -disorders	304.53	ICD-9	Hallucinogen dependence, in remission
Substance-use -disorders	304.6	ICD-9	Other specified drug dependence, unspecified
Substance-use -disorders	304.61	ICD-9	Other specified drug dependence, continuous
Substance-use -disorders	304.62	ICD-9	Other specified drug dependence, episodic
Substance-use -disorders	304.63	ICD-9	Other specified drug dependence, in remission
Substance-use -disorders	304.7	ICD-9	Combinations of opioid type drug with any other drug dependence, unspecified
Substance-use -disorders	304.71	ICD-9	Combinations of opioid type drug with any other drug dependence, continuous
Substance-use -disorders	304.72	ICD-9	Combinations of opioid type drug with any other drug dependence, episodic
Substance-use -disorders	304.73	ICD-9	Combinations of opioid type drug with any other drug dependence, in remission
Substance-use -disorders	304.8	ICD-9	Combinations of drug dependence excluding opioid type drug, unspecified
Substance-use -disorders	304.81	ICD-9	Combinations of drug dependence excluding opioid type drug, continuous
Substance-use -disorders	304.82	ICD-9	Combinations of drug dependence excluding opioid type drug, episodic
Substance-use -disorders	304.83	ICD-9	Combinations of drug dependence excluding opioid type drug, in remission
Substance-use -disorders	304.9	ICD-9	Unspecified drug dependence, unspecified

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	304.91	ICD-9	Unspecified drug dependence, continuous
Substance-use -disorders	304.92	ICD-9	Unspecified drug dependence, episodic
Substance-use -disorders	304.93	ICD-9	Unspecified drug dependence, in remission
Substance-use -disorders	305	ICD-9	Alcohol abuse, unspecified
Substance-use -disorders	305.01	ICD-9	Alcohol abuse, continuous
Substance-use -disorders	305.02	ICD-9	Alcohol abuse, episodic
Substance-use -disorders	305.03	ICD-9	Alcohol abuse, in remission
Substance-use -disorders	305.1	ICD-9	Tobacco use disorder
Substance-use -disorders	305.2	ICD-9	Cannabis abuse, unspecified
Substance-use -disorders	305.21	ICD-9	Cannabis abuse, continuous
Substance-use -disorders	305.22	ICD-9	Cannabis abuse, episodic
Substance-use -disorders	305.23	ICD-9	Cannabis abuse, in remission
Substance-use -disorders	305.3	ICD-9	Hallucinogen abuse, unspecified
Substance-use -disorders	305.31	ICD-9	Hallucinogen abuse, continuous
Substance-use -disorders	305.32	ICD-9	Hallucinogen abuse, episodic
Substance-use -disorders	305.33	ICD-9	Hallucinogen abuse, in remission
Substance-use -disorders	305.4	ICD-9	Sedative, hypnotic or anxiolytic abuse, unspecified
Substance-use -disorders	305.41	ICD-9	Sedative, hypnotic or anxiolytic abuse, continuous
Substance-use -disorders	305.42	ICD-9	Sedative, hypnotic or anxiolytic abuse, episodic
Substance-use -disorders	305.43	ICD-9	Sedative, hypnotic or anxiolytic abuse, in remission
Substance-use -disorders	305.5	ICD-9	Opioid abuse, unspecified
Substance-use -disorders	305.51	ICD-9	Opioid abuse, continuous
Substance-use -disorders	305.52	ICD-9	Opioid abuse, episodic
Substance-use -disorders	305.53	ICD-9	Opioid abuse, in remission
Substance-use -disorders	305.6	ICD-9	Cocaine abuse, unspecified

For peer review only

Online supplementary material

Substance-use -disorders	305.61	ICD-9	Cocaine abuse, continuous
Substance-use -disorders	305.62	ICD-9	Cocaine abuse, episodic
Substance-use -disorders	305.63	ICD-9	Cocaine abuse, in remission
Substance-use -disorders	305.7	ICD-9	Amphetamine or related acting sympathomimetic abuse, unspecified
Substance-use -disorders	305.71	ICD-9	Amphetamine or related acting sympathomimetic abuse, continuous
Substance-use -disorders	305.72	ICD-9	Amphetamine or related acting sympathomimetic abuse, episodic
Substance-use -disorders	305.73	ICD-9	Amphetamine or related acting sympathomimetic abuse, in remission
Substance-use -disorders	305.8	ICD-9	Antidepressant type abuse, unspecified
Substance-use -disorders	305.81	ICD-9	Antidepressant type abuse, continuous
Substance-use -disorders	305.82	ICD-9	Antidepressant type abuse, episodic
Substance-use -disorders	305.83	ICD-9	Antidepressant type abuse, in remission
Substance-use -disorders	305.9	ICD-9	Other, mixed, or unspecified drug abuse, unspecified
Substance-use -disorders	305.91	ICD-9	Other, mixed, or unspecified drug abuse, continuous
Substance-use -disorders	305.92	ICD-9	Other, mixed, or unspecified drug abuse, episodic
Substance-use -disorders	305.93	ICD-9	Other, mixed, or unspecified drug abuse, in remission
Substance-use -disorders	F10.10	ICD-10	Alcohol abuse, uncomplicated
Substance-use -disorders	F10.11	ICD-10	Alcohol abuse, in remission
Substance-use -disorders	F10.120	ICD-10	Alcohol abuse with intoxication, uncomplicated
Substance-use -disorders	F10.121	ICD-10	Alcohol abuse with intoxication delirium
Substance-use -disorders	F10.129	ICD-10	Alcohol abuse with intoxication, unspecified
Substance-use -disorders	F10.14	ICD-10	Alcohol abuse with alcohol-induced mood disorder
Substance-use -disorders	F10.150	ICD-10	Alcohol abuse with alcohol-induced psychotic disorder with delusions
Substance-use -disorders	F10.151	ICD-10	Alcohol abuse with alcohol-induced psychotic disorder with hallucinations
Substance-use -disorders	F10.159	ICD-10	Alcohol abuse with alcohol-induced psychotic disorder, unspecified
Substance-use -disorders	F10.180	ICD-10	Alcohol abuse with alcohol-induced anxiety disorder

For peer review only

Online supplementary material

Substance-use -disorders	F10.181	ICD-10	Alcohol abuse with alcohol-induced sexual dysfunction
Substance-use -disorders	F10.182	ICD-10	Alcohol abuse with alcohol-induced sleep disorder
Substance-use -disorders	F10.188	ICD-10	Alcohol abuse with other alcohol-induced disorder
Substance-use -disorders	F10.19	ICD-10	Alcohol abuse with unspecified alcohol-induced disorder
Substance-use -disorders	F10.20	ICD-10	Alcohol dependence, uncomplicated
Substance-use -disorders	F10.21	ICD-10	Alcohol dependence, in remission
Substance-use -disorders	F10.220	ICD-10	Alcohol dependence with intoxication, uncomplicated
Substance-use -disorders	F10.221	ICD-10	Alcohol dependence with intoxication delirium
Substance-use -disorders	F10.229	ICD-10	Alcohol dependence with intoxication, unspecified
Substance-use -disorders	F10.230	ICD-10	Alcohol dependence with withdrawal, uncomplicated
Substance-use -disorders	F10.231	ICD-10	Alcohol dependence with withdrawal delirium
Substance-use -disorders	F10.232	ICD-10	Alcohol dependence with withdrawal with perceptual disturbance
Substance-use -disorders	F10.239	ICD-10	Alcohol dependence with withdrawal, unspecified
Substance-use -disorders	F10.24	ICD-10	Alcohol dependence with alcohol-induced mood disorder
Substance-use -disorders	F10.250	ICD-10	Alcohol dependence with alcohol-induced psychotic disorder with delusions
Substance-use -disorders	F10.251	ICD-10	Alcohol dependence with alcohol-induced psychotic disorder with hallucinations
Substance-use -disorders	F10.259	ICD-10	Alcohol dependence with alcohol-induced psychotic disorder, unspecified
Substance-use -disorders	F10.26	ICD-10	Alcohol dependence with alcohol-induced persisting amnestic disorder
Substance-use -disorders	F10.27	ICD-10	Alcohol dependence with alcohol-induced persisting dementia
Substance-use -disorders	F10.280	ICD-10	Alcohol dependence with alcohol-induced anxiety disorder
Substance-use -disorders	F10.281	ICD-10	Alcohol dependence with alcohol-induced sexual dysfunction
Substance-use -disorders	F10.282	ICD-10	Alcohol dependence with alcohol-induced sleep disorder
Substance-use -disorders	F10.288	ICD-10	Alcohol dependence with other alcohol-induced disorder
Substance-use -disorders	F10.29	ICD-10	Alcohol dependence with unspecified alcohol-induced disorder
Substance-use -disorders	F10.920	ICD-10	Alcohol use, unspecified with intoxication, uncomplicated

For peer review only

Online supplementary material

Substance-use -disorders	F10.921	ICD-10	Alcohol use, unspecified with intoxication delirium
Substance-use -disorders	F10.929	ICD-10	Alcohol use, unspecified with intoxication, unspecified
Substance-use -disorders	F10.94	ICD-10	Alcohol use, unspecified with alcohol-induced mood disorder
Substance-use -disorders	F10.950	ICD-10	Alcohol use, unspecified with alcohol-induced psychotic disorder with delusions
Substance-use -disorders	F10.951	ICD-10	Alcohol use, unspecified with alcohol-induced psychotic disorder with hallucinations
Substance-use -disorders	F10.959	ICD-10	Alcohol use, unspecified with alcohol-induced psychotic disorder, unspecified
Substance-use -disorders	F10.96	ICD-10	Alcohol use, unspecified with alcohol-induced persisting amnestic disorder
Substance-use -disorders	F10.97	ICD-10	Alcohol use, unspecified with alcohol-induced persisting dementia
Substance-use -disorders	F10.980	ICD-10	Alcohol use, unspecified with alcohol-induced anxiety disorder
Substance-use -disorders	F10.981	ICD-10	Alcohol use, unspecified with alcohol-induced sexual dysfunction
Substance-use -disorders	F10.982	ICD-10	Alcohol use, unspecified with alcohol-induced sleep disorder
Substance-use -disorders	F10.988	ICD-10	Alcohol use, unspecified with other alcohol-induced disorder
Substance-use -disorders	F10.99	ICD-10	Alcohol use, unspecified with unspecified alcohol-induced disorder
Substance-use -disorders	F11.10	ICD-10	Opioid abuse, uncomplicated
Substance-use -disorders	F11.11	ICD-10	Opioid abuse, in remission
Substance-use -disorders	F11.120	ICD-10	Opioid abuse with intoxication, uncomplicated
Substance-use -disorders	F11.121	ICD-10	Opioid abuse with intoxication delirium
Substance-use -disorders	F11.122	ICD-10	Opioid abuse with intoxication with perceptual disturbance
Substance-use -disorders	F11.129	ICD-10	Opioid abuse with intoxication, unspecified
Substance-use -disorders	F11.14	ICD-10	Opioid abuse with opioid-induced mood disorder
Substance-use -disorders	F11.150	ICD-10	Opioid abuse with opioid-induced psychotic disorder with delusions
Substance-use -disorders	F11.151	ICD-10	Opioid abuse with opioid-induced psychotic disorder with hallucinations
Substance-use -disorders	F11.159	ICD-10	Opioid abuse with opioid-induced psychotic disorder, unspecified
Substance-use -disorders	F11.181	ICD-10	Opioid abuse with opioid-induced sexual dysfunction
Substance-use -disorders	F11.182	ICD-10	Opioid abuse with opioid-induced sleep disorder

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F11.188	ICD-10	Opioid abuse with other opioid-induced disorder
Substance-use -disorders	F11.19	ICD-10	Opioid abuse with unspecified opioid-induced disorder
Substance-use -disorders	F11.20	ICD-10	Opioid dependence, uncomplicated
Substance-use -disorders	F11.21	ICD-10	Opioid dependence, in remission
Substance-use -disorders	F11.220	ICD-10	Opioid dependence with intoxication, uncomplicated
Substance-use -disorders	F11.221	ICD-10	Opioid dependence with intoxication delirium
Substance-use -disorders	F11.222	ICD-10	Opioid dependence with intoxication with perceptual disturbance
Substance-use -disorders	F11.229	ICD-10	Opioid dependence with intoxication, unspecified
Substance-use -disorders	F11.23	ICD-10	Opioid dependence with withdrawal
Substance-use -disorders	F11.24	ICD-10	Opioid dependence with opioid-induced mood disorder
Substance-use -disorders	F11.250	ICD-10	Opioid dependence with opioid-induced psychotic disorder with delusions
Substance-use -disorders	F11.251	ICD-10	Opioid dependence with opioid-induced psychotic disorder with hallucinations
Substance-use -disorders	F11.259	ICD-10	Opioid dependence with opioid-induced psychotic disorder, unspecified
Substance-use -disorders	F11.281	ICD-10	Opioid dependence with opioid-induced sexual dysfunction
Substance-use -disorders	F11.282	ICD-10	Opioid dependence with opioid-induced sleep disorder
Substance-use -disorders	F11.288	ICD-10	Opioid dependence with other opioid-induced disorder
Substance-use -disorders	F11.29	ICD-10	Opioid dependence with unspecified opioid-induced disorder
Substance-use -disorders	F11.90	ICD-10	Opioid use, unspecified, uncomplicated
Substance-use -disorders	F11.920	ICD-10	Opioid use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F11.921	ICD-10	Opioid use, unspecified with intoxication delirium
Substance-use -disorders	F11.922	ICD-10	Opioid use, unspecified with intoxication with perceptual disturbance
Substance-use -disorders	F11.929	ICD-10	Opioid use, unspecified with intoxication, unspecified
Substance-use -disorders	F11.93	ICD-10	Opioid use, unspecified with withdrawal
Substance-use -disorders	F11.94	ICD-10	Opioid use, unspecified with opioid-induced mood disorder
Substance-use -disorders	F11.950	ICD-10	Opioid use, unspecified with opioid-induced psychotic disorder with delusions

For peer review only

Online supplementary material

Substance-use -disorders	F11.951	ICD-10	Opioid use, unspecified with opioid-induced psychotic disorder with hallucinations
Substance-use -disorders	F11.959	ICD-10	Opioid use, unspecified with opioid-induced psychotic disorder, unspecified
Substance-use -disorders	F11.981	ICD-10	Opioid use, unspecified with opioid-induced sexual dysfunction
Substance-use -disorders	F11.982	ICD-10	Opioid use, unspecified with opioid-induced sleep disorder
Substance-use -disorders	F11.988	ICD-10	Opioid use, unspecified with other opioid-induced disorder
Substance-use -disorders	F11.99	ICD-10	Opioid use, unspecified with unspecified opioid-induced disorder
Substance-use -disorders	F12.10	ICD-10	Cannabis abuse, uncomplicated
Substance-use -disorders	F12.11	ICD-10	Cannabis abuse, in remission
Substance-use -disorders	F12.120	ICD-10	Cannabis abuse with intoxication, uncomplicated
Substance-use -disorders	F12.121	ICD-10	Cannabis abuse with intoxication delirium
Substance-use -disorders	F12.122	ICD-10	Cannabis abuse with intoxication with perceptual disturbance
Substance-use -disorders	F12.129	ICD-10	Cannabis abuse with intoxication, unspecified
Substance-use -disorders	F12.150	ICD-10	Cannabis abuse with psychotic disorder with delusions
Substance-use -disorders	F12.151	ICD-10	Cannabis abuse with psychotic disorder with hallucinations
Substance-use -disorders	F12.159	ICD-10	Cannabis abuse with psychotic disorder, unspecified
Substance-use -disorders	F12.180	ICD-10	Cannabis abuse with cannabis-induced anxiety disorder
Substance-use -disorders	F12.188	ICD-10	Cannabis abuse with other cannabis-induced disorder
Substance-use -disorders	F12.19	ICD-10	Cannabis abuse with unspecified cannabis-induced disorder
Substance-use -disorders	F12.20	ICD-10	Cannabis dependence, uncomplicated
Substance-use -disorders	F12.21	ICD-10	Cannabis dependence, in remission
Substance-use -disorders	F12.220	ICD-10	Cannabis dependence with intoxication, uncomplicated
Substance-use -disorders	F12.221	ICD-10	Cannabis dependence with intoxication delirium
Substance-use -disorders	F12.222	ICD-10	Cannabis dependence with intoxication with perceptual disturbance
Substance-use -disorders	F12.229	ICD-10	Cannabis dependence with intoxication, unspecified
Substance-use -disorders	F12.250	ICD-10	Cannabis dependence with psychotic disorder with delusions

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F12.251	ICD-10	Cannabis dependence with psychotic disorder with hallucinations
Substance-use -disorders	F12.259	ICD-10	Cannabis dependence with psychotic disorder, unspecified
Substance-use -disorders	F12.280	ICD-10	Cannabis dependence with cannabis-induced anxiety disorder
Substance-use -disorders	F12.288	ICD-10	Cannabis dependence with other cannabis-induced disorder
Substance-use -disorders	F12.29	ICD-10	Cannabis dependence with unspecified cannabis-induced disorder
Substance-use -disorders	F12.90	ICD-10	Cannabis use, unspecified, uncomplicated
Substance-use -disorders	F12.920	ICD-10	Cannabis use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F12.921	ICD-10	Cannabis use, unspecified with intoxication delirium
Substance-use -disorders	F12.922	ICD-10	Cannabis use, unspecified with intoxication with perceptual disturbance
Substance-use -disorders	F12.929	ICD-10	Cannabis use, unspecified with intoxication, unspecified
Substance-use -disorders	F12.950	ICD-10	Cannabis use, unspecified with psychotic disorder with delusions
Substance-use -disorders	F12.951	ICD-10	Cannabis use, unspecified with psychotic disorder with hallucinations
Substance-use -disorders	F12.959	ICD-10	Cannabis use, unspecified with psychotic disorder, unspecified
Substance-use -disorders	F12.980	ICD-10	Cannabis use, unspecified with anxiety disorder
Substance-use -disorders	F12.988	ICD-10	Cannabis use, unspecified with other cannabis-induced disorder
Substance-use -disorders	F12.99	ICD-10	Cannabis use, unspecified with unspecified cannabis-induced disorder
Substance-use -disorders	F13.10	ICD-10	Sedative, hypnotic or anxiolytic abuse, uncomplicated
Substance-use -disorders	F13.11	ICD-10	Sedative, hypnotic or anxiolytic abuse, in remission
Substance-use -disorders	F13.120	ICD-10	Sedative, hypnotic or anxiolytic abuse with intoxication, uncomplicated
Substance-use -disorders	F13.121	ICD-10	Sedative, hypnotic or anxiolytic abuse with intoxication delirium
Substance-use -disorders	F13.129	ICD-10	Sedative, hypnotic or anxiolytic abuse with intoxication, unspecified
Substance-use -disorders	F13.14	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced mood disorder
Substance-use -disorders	F13.150	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced psychotic disorder with delusions
Substance-use -disorders	F13.151	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced psychotic disorder with hallucinations

For peer review only

Online supplementary material

Substance-use -disorders	F13.159	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced psychotic disorder, unspecified
Substance-use -disorders	F13.180	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced anxiety disorder
Substance-use -disorders	F13.181	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced sexual dysfunction
Substance-use -disorders	F13.182	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced sleep disorder
Substance-use -disorders	F13.188	ICD-10	Sedative, hypnotic or anxiolytic abuse with other sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F13.19	ICD-10	Sedative, hypnotic or anxiolytic abuse with unspecified sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F13.20	ICD-10	Sedative, hypnotic or anxiolytic dependence, uncomplicated
Substance-use -disorders	F13.21	ICD-10	Sedative, hypnotic or anxiolytic dependence, in remission
Substance-use -disorders	F13.220	ICD-10	Sedative, hypnotic or anxiolytic dependence with intoxication, uncomplicated
Substance-use -disorders	F13.221	ICD-10	Sedative, hypnotic or anxiolytic dependence with intoxication delirium
Substance-use -disorders	F13.229	ICD-10	Sedative, hypnotic or anxiolytic dependence with intoxication, unspecified
Substance-use -disorders	F13.230	ICD-10	Sedative, hypnotic or anxiolytic dependence with withdrawal, uncomplicated
Substance-use -disorders	F13.231	ICD-10	Sedative, hypnotic or anxiolytic dependence with withdrawal delirium
Substance-use -disorders	F13.232	ICD-10	Sedative, hypnotic or anxiolytic dependence with withdrawal with perceptual disturbance
Substance-use -disorders	F13.239	ICD-10	Sedative, hypnotic or anxiolytic dependence with withdrawal, unspecified
Substance-use -disorders	F13.24	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced mood disorder
Substance-use -disorders	F13.250	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced psychotic disorder with delusions
Substance-use -disorders	F13.251	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced psychotic disorder with hallucinations
Substance-use -disorders	F13.259	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced psychotic disorder, unspecified
Substance-use -disorders	F13.26	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced persisting amnestic disorder
Substance-use -disorders	F13.27	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced persisting dementia
Substance-use -disorders	F13.280	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced anxiety disorder

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F13.281	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced sexual dysfunction
Substance-use -disorders	F13.282	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced sleep disorder
Substance-use -disorders	F13.288	ICD-10	Sedative, hypnotic or anxiolytic dependence with other sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F13.29	ICD-10	Sedative, hypnotic or anxiolytic dependence with unspecified sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F13.90	ICD-10	Sedative, hypnotic, or anxiolytic use, unspecified, uncomplicated
Substance-use -disorders	F13.920	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F13.921	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with intoxication delirium
Substance-use -disorders	F13.929	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with intoxication, unspecified
Substance-use -disorders	F13.930	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with withdrawal, uncomplicated
Substance-use -disorders	F13.931	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with withdrawal delirium
Substance-use -disorders	F13.932	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with withdrawal with perceptual disturbances
Substance-use -disorders	F13.939	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with withdrawal, unspecified
Substance-use -disorders	F13.94	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced mood disorder
Substance-use -disorders	F13.950	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced psychotic disorder with delusions
Substance-use -disorders	F13.951	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced psychotic disorder with hallucinations
Substance-use -disorders	F13.959	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced psychotic disorder, unspecified
Substance-use -disorders	F13.96	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced persisting amnestic disorder
Substance-use -disorders	F13.97	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced persisting dementia
Substance-use -disorders	F13.980	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced anxiety disorder
Substance-use -disorders	F13.981	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced sexual dysfunction
Substance-use -disorders	F13.982	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced sleep disorder

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F13.988	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with other sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F13.99	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with unspecified sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F14.10	ICD-10	Cocaine abuse, uncomplicated
Substance-use -disorders	F14.11	ICD-10	Cocaine abuse, in remission
Substance-use -disorders	F14.120	ICD-10	Cocaine abuse with intoxication, uncomplicated
Substance-use -disorders	F14.121	ICD-10	Cocaine abuse with intoxication with delirium
Substance-use -disorders	F14.122	ICD-10	Cocaine abuse with intoxication with perceptual disturbance
Substance-use -disorders	F14.129	ICD-10	Cocaine abuse with intoxication, unspecified
Substance-use -disorders	F14.14	ICD-10	Cocaine abuse with cocaine-induced mood disorder
Substance-use -disorders	F14.150	ICD-10	Cocaine abuse with cocaine-induced psychotic disorder with delusions
Substance-use -disorders	F14.151	ICD-10	Cocaine abuse with cocaine-induced psychotic disorder with hallucinations
Substance-use -disorders	F14.159	ICD-10	Cocaine abuse with cocaine-induced psychotic disorder, unspecified
Substance-use -disorders	F14.180	ICD-10	Cocaine abuse with cocaine-induced anxiety disorder
Substance-use -disorders	F14.181	ICD-10	Cocaine abuse with cocaine-induced sexual dysfunction
Substance-use -disorders	F14.182	ICD-10	Cocaine abuse with cocaine-induced sleep disorder
Substance-use -disorders	F14.188	ICD-10	Cocaine abuse with other cocaine-induced disorder
Substance-use -disorders	F14.19	ICD-10	Cocaine abuse with unspecified cocaine-induced disorder
Substance-use -disorders	F14.20	ICD-10	Cocaine dependence, uncomplicated
Substance-use -disorders	F14.21	ICD-10	Cocaine dependence, in remission
Substance-use -disorders	F14.220	ICD-10	Cocaine dependence with intoxication, uncomplicated
Substance-use -disorders	F14.221	ICD-10	Cocaine dependence with intoxication delirium
Substance-use -disorders	F14.222	ICD-10	Cocaine dependence with intoxication with perceptual disturbance
Substance-use -disorders	F14.229	ICD-10	Cocaine dependence with intoxication, unspecified
Substance-use -disorders	F14.23	ICD-10	Cocaine dependence with withdrawal
Substance-use -disorders	F14.24	ICD-10	Cocaine dependence with cocaine-induced mood disorder

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F14.250	ICD-10	Cocaine dependence with cocaine-induced psychotic disorder with delusions
Substance-use -disorders	F14.251	ICD-10	Cocaine dependence with cocaine-induced psychotic disorder with hallucinations
Substance-use -disorders	F14.259	ICD-10	Cocaine dependence with cocaine-induced psychotic disorder, unspecified
Substance-use -disorders	F14.280	ICD-10	Cocaine dependence with cocaine-induced anxiety disorder
Substance-use -disorders	F14.281	ICD-10	Cocaine dependence with cocaine-induced sexual dysfunction
Substance-use -disorders	F14.282	ICD-10	Cocaine dependence with cocaine-induced sleep disorder
Substance-use -disorders	F14.288	ICD-10	Cocaine dependence with other cocaine-induced disorder
Substance-use -disorders	F14.29	ICD-10	Cocaine dependence with unspecified cocaine-induced disorder
Substance-use -disorders	F14.90	ICD-10	Cocaine use, unspecified, uncomplicated
Substance-use -disorders	F14.920	ICD-10	Cocaine use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F14.921	ICD-10	Cocaine use, unspecified with intoxication delirium
Substance-use -disorders	F14.922	ICD-10	Cocaine use, unspecified with intoxication with perceptual disturbance
Substance-use -disorders	F14.929	ICD-10	Cocaine use, unspecified with intoxication, unspecified
Substance-use -disorders	F14.94	ICD-10	Cocaine use, unspecified with cocaine-induced mood disorder
Substance-use -disorders	F14.950	ICD-10	Cocaine use, unspecified with cocaine-induced psychotic disorder with delusions
Substance-use -disorders	F14.951	ICD-10	Cocaine use, unspecified with cocaine-induced psychotic disorder with hallucinations
Substance-use -disorders	F14.959	ICD-10	Cocaine use, unspecified with cocaine-induced psychotic disorder, unspecified
Substance-use -disorders	F14.980	ICD-10	Cocaine use, unspecified with cocaine-induced anxiety disorder
Substance-use -disorders	F14.981	ICD-10	Cocaine use, unspecified with cocaine-induced sexual dysfunction
Substance-use -disorders	F14.982	ICD-10	Cocaine use, unspecified with cocaine-induced sleep disorder
Substance-use -disorders	F14.988	ICD-10	Cocaine use, unspecified with other cocaine-induced disorder
Substance-use -disorders	F14.99	ICD-10	Cocaine use, unspecified with unspecified cocaine-induced disorder
Substance-use -disorders	F15.10	ICD-10	Other stimulant abuse, uncomplicated
Substance-use -disorders	F15.11	ICD-10	Other stimulant abuse, in remission
Substance-use -disorders	F15.120	ICD-10	Other stimulant abuse with intoxication, uncomplicated

For peer review only

Online supplementary material

Substance-use -disorders	F15.121	ICD-10	Other stimulant abuse with intoxication delirium
Substance-use -disorders	F15.122	ICD-10	Other stimulant abuse with intoxication with perceptual disturbance
Substance-use -disorders	F15.129	ICD-10	Other stimulant abuse with intoxication, unspecified
Substance-use -disorders	F15.14	ICD-10	Other stimulant abuse with stimulant-induced mood disorder
Substance-use -disorders	F15.150	ICD-10	Other stimulant abuse with stimulant-induced psychotic disorder with delusions
Substance-use -disorders	F15.151	ICD-10	Other stimulant abuse with stimulant-induced psychotic disorder with hallucinations
Substance-use -disorders	F15.159	ICD-10	Other stimulant abuse with stimulant-induced psychotic disorder, unspecified
Substance-use -disorders	F15.180	ICD-10	Other stimulant abuse with stimulant-induced anxiety disorder
Substance-use -disorders	F15.181	ICD-10	Other stimulant abuse with stimulant-induced sexual dysfunction
Substance-use -disorders	F15.182	ICD-10	Other stimulant abuse with stimulant-induced sleep disorder
Substance-use -disorders	F15.188	ICD-10	Other stimulant abuse with other stimulant-induced disorder
Substance-use -disorders	F15.19	ICD-10	Other stimulant abuse with unspecified stimulant-induced disorder
Substance-use -disorders	F15.20	ICD-10	Other stimulant dependence, uncomplicated
Substance-use -disorders	F15.21	ICD-10	Other stimulant dependence, in remission
Substance-use -disorders	F15.220	ICD-10	Other stimulant dependence with intoxication, uncomplicated
Substance-use -disorders	F15.221	ICD-10	Other stimulant dependence with intoxication delirium
Substance-use -disorders	F15.222	ICD-10	Other stimulant dependence with intoxication with perceptual disturbance
Substance-use -disorders	F15.229	ICD-10	Other stimulant dependence with intoxication, unspecified
Substance-use -disorders	F15.23	ICD-10	Other stimulant dependence with withdrawal
Substance-use -disorders	F15.24	ICD-10	Other stimulant dependence with stimulant-induced mood disorder
Substance-use -disorders	F15.250	ICD-10	Other stimulant dependence with stimulant-induced psychotic disorder with delusions
Substance-use -disorders	F15.251	ICD-10	Other stimulant dependence with stimulant-induced psychotic disorder with hallucinations
Substance-use -disorders	F15.259	ICD-10	Other stimulant dependence with stimulant-induced psychotic disorder, unspecified
Substance-use -disorders	F15.280	ICD-10	Other stimulant dependence with stimulant-induced anxiety disorder
Substance-use -disorders	F15.281	ICD-10	Other stimulant dependence with stimulant-induced sexual dysfunction

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F15.282	ICD-10	Other stimulant dependence with stimulant-induced sleep disorder
Substance-use -disorders	F15.288	ICD-10	Other stimulant dependence with other stimulant-induced disorder
Substance-use -disorders	F15.29	ICD-10	Other stimulant dependence with unspecified stimulant-induced disorder
Substance-use -disorders	F15.90	ICD-10	Other stimulant use, unspecified, uncomplicated
Substance-use -disorders	F15.920	ICD-10	Other stimulant use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F15.921	ICD-10	Other stimulant use, unspecified with intoxication delirium
Substance-use -disorders	F15.922	ICD-10	Other stimulant use, unspecified with intoxication with perceptual disturbance
Substance-use -disorders	F15.929	ICD-10	Other stimulant use, unspecified with intoxication, unspecified
Substance-use -disorders	F15.93	ICD-10	Other stimulant use, unspecified with withdrawal
Substance-use -disorders	F15.94	ICD-10	Other stimulant use, unspecified with stimulant-induced mood disorder
Substance-use -disorders	F15.950	ICD-10	Other stimulant use, unspecified with stimulant-induced psychotic disorder with delusions
Substance-use -disorders	F15.951	ICD-10	Other stimulant use, unspecified with stimulant-induced psychotic disorder with hallucinations
Substance-use -disorders	F15.959	ICD-10	Other stimulant use, unspecified with stimulant-induced psychotic disorder, unspecified
Substance-use -disorders	F15.980	ICD-10	Other stimulant use, unspecified with stimulant-induced anxiety disorder
Substance-use -disorders	F15.981	ICD-10	Other stimulant use, unspecified with stimulant-induced sexual dysfunction
Substance-use -disorders	F15.982	ICD-10	Other stimulant use, unspecified with stimulant-induced sleep disorder
Substance-use -disorders	F15.988	ICD-10	Other stimulant use, unspecified with other stimulant-induced disorder
Substance-use -disorders	F15.99	ICD-10	Other stimulant use, unspecified with unspecified stimulant-induced disorder
Substance-use -disorders	F16.10	ICD-10	Hallucinogen abuse, uncomplicated
Substance-use -disorders	F16.11	ICD-10	Hallucinogen abuse, in remission
Substance-use -disorders	F16.120	ICD-10	Hallucinogen abuse with intoxication, uncomplicated
Substance-use -disorders	F16.121	ICD-10	Hallucinogen abuse with intoxication with delirium
Substance-use -disorders	F16.122	ICD-10	Hallucinogen abuse with intoxication with perceptual disturbance
Substance-use -disorders	F16.129	ICD-10	Hallucinogen abuse with intoxication, unspecified
Substance-use -disorders	F16.14	ICD-10	Hallucinogen abuse with hallucinogen-induced mood disorder

For peer review only

Online supplementary material

Substance-use -disorders	F16.150	ICD-10	Hallucinogen abuse with hallucinogen-induced psychotic disorder with delusions
Substance-use -disorders	F16.151	ICD-10	Hallucinogen abuse with hallucinogen-induced psychotic disorder with hallucinations
Substance-use -disorders	F16.159	ICD-10	Hallucinogen abuse with hallucinogen-induced psychotic disorder, unspecified
Substance-use -disorders	F16.180	ICD-10	Hallucinogen abuse with hallucinogen-induced anxiety disorder
Substance-use -disorders	F16.183	ICD-10	Hallucinogen abuse with hallucinogen persisting perception disorder (flashbacks)
Substance-use -disorders	F16.188	ICD-10	Hallucinogen abuse with other hallucinogen-induced disorder
Substance-use -disorders	F16.19	ICD-10	Hallucinogen abuse with unspecified hallucinogen-induced disorder
Substance-use -disorders	F16.20	ICD-10	Hallucinogen dependence, uncomplicated
Substance-use -disorders	F16.21	ICD-10	Hallucinogen dependence, in remission
Substance-use -disorders	F16.220	ICD-10	Hallucinogen dependence with intoxication, uncomplicated
Substance-use -disorders	F16.221	ICD-10	Hallucinogen dependence with intoxication with delirium
Substance-use -disorders	F16.229	ICD-10	Hallucinogen dependence with intoxication, unspecified
Substance-use -disorders	F16.24	ICD-10	Hallucinogen dependence with hallucinogen-induced mood disorder
Substance-use -disorders	F16.250	ICD-10	Hallucinogen dependence with hallucinogen-induced psychotic disorder with delusions
Substance-use -disorders	F16.251	ICD-10	Hallucinogen dependence with hallucinogen-induced psychotic disorder with hallucinations
Substance-use -disorders	F16.259	ICD-10	Hallucinogen dependence with hallucinogen-induced psychotic disorder, unspecified
Substance-use -disorders	F16.280	ICD-10	Hallucinogen dependence with hallucinogen-induced anxiety disorder
Substance-use -disorders	F16.283	ICD-10	Hallucinogen dependence with hallucinogen persisting perception disorder (flashbacks)
Substance-use -disorders	F16.288	ICD-10	Hallucinogen dependence with other hallucinogen-induced disorder
Substance-use -disorders	F16.29	ICD-10	Hallucinogen dependence with unspecified hallucinogen-induced disorder
Substance-use -disorders	F16.90	ICD-10	Hallucinogen use, unspecified, uncomplicated
Substance-use -disorders	F16.920	ICD-10	Hallucinogen use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F16.921	ICD-10	Hallucinogen use, unspecified with intoxication with delirium
Substance-use -disorders	F16.929	ICD-10	Hallucinogen use, unspecified with intoxication, unspecified
Substance-use -disorders	F16.94	ICD-10	Hallucinogen use, unspecified with hallucinogen-induced mood disorder

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F16.950	ICD-10	Hallucinogen use, unspecified with hallucinogen-induced psychotic disorder with delusions
Substance-use -disorders	F16.951	ICD-10	Hallucinogen use, unspecified with hallucinogen-induced psychotic disorder with hallucinations
Substance-use -disorders	F16.959	ICD-10	Hallucinogen use, unspecified with hallucinogen-induced psychotic disorder, unspecified
Substance-use -disorders	F16.980	ICD-10	Hallucinogen use, unspecified with hallucinogen-induced anxiety disorder
Substance-use -disorders	F16.983	ICD-10	Hallucinogen use, unspecified with hallucinogen persisting perception disorder (flashbacks)
Substance-use -disorders	F16.988	ICD-10	Hallucinogen use, unspecified with other hallucinogen-induced disorder
Substance-use -disorders	F16.99	ICD-10	Hallucinogen use, unspecified with unspecified hallucinogen-induced disorder
Substance-use -disorders	F17.200	ICD-10	Nicotine dependence, unspecified, uncomplicated
Substance-use -disorders	F17.201	ICD-10	Nicotine dependence, unspecified, in remission
Substance-use -disorders	F17.203	ICD-10	Nicotine dependence unspecified, with withdrawal
Substance-use -disorders	F17.208	ICD-10	Nicotine dependence, unspecified, with other nicotine-induced disorders
Substance-use -disorders	F17.209	ICD-10	Nicotine dependence, unspecified, with unspecified nicotine-induced disorders
Substance-use -disorders	F17.210	ICD-10	Nicotine dependence, cigarettes, uncomplicated
Substance-use -disorders	F17.211	ICD-10	Nicotine dependence, cigarettes, in remission
Substance-use -disorders	F17.213	ICD-10	Nicotine dependence, cigarettes, with withdrawal
Substance-use -disorders	F17.218	ICD-10	Nicotine dependence, cigarettes, with other nicotine-induced disorders
Substance-use -disorders	F17.219	ICD-10	Nicotine dependence, cigarettes, with unspecified nicotine-induced disorders
Substance-use -disorders	F17.220	ICD-10	Nicotine dependence, chewing tobacco, uncomplicated
Substance-use -disorders	F17.221	ICD-10	Nicotine dependence, chewing tobacco, in remission
Substance-use -disorders	F17.223	ICD-10	Nicotine dependence, chewing tobacco, with withdrawal
Substance-use -disorders	F17.228	ICD-10	Nicotine dependence, chewing tobacco, with other nicotine-induced disorders
Substance-use -disorders	F17.229	ICD-10	Nicotine dependence, chewing tobacco, with unspecified nicotine-induced disorders
Substance-use -disorders	F17.290	ICD-10	Nicotine dependence, other tobacco product, uncomplicated
Substance-use -disorders	F17.291	ICD-10	Nicotine dependence, other tobacco product, in remission
Substance-use -disorders	F17.293	ICD-10	Nicotine dependence, other tobacco product, with withdrawal

For peer review only

Online supplementary material

Substance-use -disorders	F17.298	ICD-10	Nicotine dependence, other tobacco product, with other nicotine-induced disorders
Substance-use -disorders	F17.299	ICD-10	Nicotine dependence, other tobacco product, with unspecified nicotine-induced disorders
Substance-use -disorders	F18.10	ICD-10	Inhalant abuse, uncomplicated
Substance-use -disorders	F18.11	ICD-10	Inhalant abuse, in remission
Substance-use -disorders	F18.120	ICD-10	Inhalant abuse with intoxication, uncomplicated
Substance-use -disorders	F18.121	ICD-10	Inhalant abuse with intoxication delirium
Substance-use -disorders	F18.129	ICD-10	Inhalant abuse with intoxication, unspecified
Substance-use -disorders	F18.14	ICD-10	Inhalant abuse with inhalant-induced mood disorder
Substance-use -disorders	F18.150	ICD-10	Inhalant abuse with inhalant-induced psychotic disorder with delusions
Substance-use -disorders	F18.151	ICD-10	Inhalant abuse with inhalant-induced psychotic disorder with hallucinations
Substance-use -disorders	F18.159	ICD-10	Inhalant abuse with inhalant-induced psychotic disorder, unspecified
Substance-use -disorders	F18.17	ICD-10	Inhalant abuse with inhalant-induced dementia
Substance-use -disorders	F18.180	ICD-10	Inhalant abuse with inhalant-induced anxiety disorder
Substance-use -disorders	F18.188	ICD-10	Inhalant abuse with other inhalant-induced disorder
Substance-use -disorders	F18.19	ICD-10	Inhalant abuse with unspecified inhalant-induced disorder
Substance-use -disorders	F18.20	ICD-10	Inhalant dependence, uncomplicated
Substance-use -disorders	F18.21	ICD-10	Inhalant dependence, in remission
Substance-use -disorders	F18.220	ICD-10	Inhalant dependence with intoxication, uncomplicated
Substance-use -disorders	F18.221	ICD-10	Inhalant dependence with intoxication delirium
Substance-use -disorders	F18.229	ICD-10	Inhalant dependence with intoxication, unspecified
Substance-use -disorders	F18.24	ICD-10	Inhalant dependence with inhalant-induced mood disorder
Substance-use -disorders	F18.250	ICD-10	Inhalant dependence with inhalant-induced psychotic disorder with delusions
Substance-use -disorders	F18.251	ICD-10	Inhalant dependence with inhalant-induced psychotic disorder with hallucinations
Substance-use -disorders	F18.259	ICD-10	Inhalant dependence with inhalant-induced psychotic disorder, unspecified
Substance-use -disorders	F18.27	ICD-10	Inhalant dependence with inhalant-induced dementia

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F18.280	ICD-10	Inhalant dependence with inhalant-induced anxiety disorder
Substance-use -disorders	F18.288	ICD-10	Inhalant dependence with other inhalant-induced disorder
Substance-use -disorders	F18.29	ICD-10	Inhalant dependence with unspecified inhalant-induced disorder
Substance-use -disorders	F18.90	ICD-10	Inhalant use, unspecified, uncomplicated
Substance-use -disorders	F18.920	ICD-10	Inhalant use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F18.921	ICD-10	Inhalant use, unspecified with intoxication with delirium
Substance-use -disorders	F18.929	ICD-10	Inhalant use, unspecified with intoxication, unspecified
Substance-use -disorders	F18.94	ICD-10	Inhalant use, unspecified with inhalant-induced mood disorder
Substance-use -disorders	F18.950	ICD-10	Inhalant use, unspecified with inhalant-induced psychotic disorder with delusions
Substance-use -disorders	F18.951	ICD-10	Inhalant use, unspecified with inhalant-induced psychotic disorder with hallucinations
Substance-use -disorders	F18.959	ICD-10	Inhalant use, unspecified with inhalant-induced psychotic disorder, unspecified
Substance-use -disorders	F18.97	ICD-10	Inhalant use, unspecified with inhalant-induced persisting dementia
Substance-use -disorders	F18.980	ICD-10	Inhalant use, unspecified with inhalant-induced anxiety disorder
Substance-use -disorders	F18.988	ICD-10	Inhalant use, unspecified with other inhalant-induced disorder
Substance-use -disorders	F18.99	ICD-10	Inhalant use, unspecified with unspecified inhalant-induced disorder
Substance-use -disorders	F19.10	ICD-10	Other psychoactive substance abuse, uncomplicated
Substance-use -disorders	F19.11	ICD-10	Other psychoactive substance abuse, in remission
Substance-use -disorders	F19.120	ICD-10	Other psychoactive substance abuse with intoxication, uncomplicated
Substance-use -disorders	F19.121	ICD-10	Other psychoactive substance abuse with intoxication delirium
Substance-use -disorders	F19.122	ICD-10	Other psychoactive substance abuse with intoxication with perceptual disturbances
Substance-use -disorders	F19.129	ICD-10	Other psychoactive substance abuse with intoxication, unspecified
Substance-use -disorders	F19.14	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced mood disorder
Substance-use -disorders	F19.150	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced psychotic disorder with delusions
Substance-use -disorders	F19.151	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced psychotic disorder with hallucinations
Substance-use -disorders	F19.159	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced psychotic disorder, unspecified

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F19.16	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced persisting amnestic disorder
Substance-use -disorders	F19.17	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced persisting dementia
Substance-use -disorders	F19.180	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced anxiety disorder
Substance-use -disorders	F19.181	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced sexual dysfunction
Substance-use -disorders	F19.182	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced sleep disorder
Substance-use -disorders	F19.188	ICD-10	Other psychoactive substance abuse with other psychoactive substance-induced disorder
Substance-use -disorders	F19.19	ICD-10	Other psychoactive substance abuse with unspecified psychoactive substance-induced disorder
Substance-use -disorders	F19.20	ICD-10	Other psychoactive substance dependence, uncomplicated
Substance-use -disorders	F19.21	ICD-10	Other psychoactive substance dependence, in remission
Substance-use -disorders	F19.220	ICD-10	Other psychoactive substance dependence with intoxication, uncomplicated
Substance-use -disorders	F19.221	ICD-10	Other psychoactive substance dependence with intoxication delirium
Substance-use -disorders	F19.222	ICD-10	Other psychoactive substance dependence with intoxication with perceptual disturbance
Substance-use -disorders	F19.229	ICD-10	Other psychoactive substance dependence with intoxication, unspecified
Substance-use -disorders	F19.230	ICD-10	Other psychoactive substance dependence with withdrawal, uncomplicated
Substance-use -disorders	F19.231	ICD-10	Other psychoactive substance dependence with withdrawal delirium
Substance-use -disorders	F19.232	ICD-10	Other psychoactive substance dependence with withdrawal with perceptual disturbance
Substance-use -disorders	F19.239	ICD-10	Other psychoactive substance dependence with withdrawal, unspecified
Substance-use -disorders	F19.24	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced mood disorder
Substance-use -disorders	F19.250	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced psychotic disorder with delusions
Substance-use -disorders	F19.251	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced psychotic disorder with hallucinations
Substance-use -disorders	F19.259	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced psychotic disorder, unspecified
Substance-use -disorders	F19.26	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced persisting amnestic disorder
Substance-use -disorders	F19.27	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced persisting dementia

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F19.280	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced anxiety disorder
Substance-use -disorders	F19.281	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced sexual dysfunction
Substance-use -disorders	F19.282	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced sleep disorder
Substance-use -disorders	F19.288	ICD-10	Other psychoactive substance dependence with other psychoactive substance-induced disorder
Substance-use -disorders	F19.29	ICD-10	Other psychoactive substance dependence with unspecified psychoactive substance-induced disorder
Substance-use -disorders	F19.90	ICD-10	Other psychoactive substance use, unspecified, uncomplicated
Substance-use -disorders	F19.920	ICD-10	Other psychoactive substance use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F19.921	ICD-10	Other psychoactive substance use, unspecified with intoxication with delirium
Substance-use -disorders	F19.922	ICD-10	Other psychoactive substance use, unspecified with intoxication with perceptual disturbance
Substance-use -disorders	F19.929	ICD-10	Other psychoactive substance use, unspecified with intoxication, unspecified
Substance-use -disorders	F19.930	ICD-10	Other psychoactive substance use, unspecified with withdrawal, uncomplicated
Substance-use -disorders	F19.931	ICD-10	Other psychoactive substance use, unspecified with withdrawal delirium
Substance-use -disorders	F19.932	ICD-10	Other psychoactive substance use, unspecified with withdrawal with perceptual disturbance
Substance-use -disorders	F19.939	ICD-10	Other psychoactive substance use, unspecified with withdrawal, unspecified
Substance-use -disorders	F19.94	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced mood disorder
Substance-use -disorders	F19.950	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced psychotic disorder with delusions
Substance-use -disorders	F19.951	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced psychotic disorder with hallucinations
Substance-use -disorders	F19.959	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced psychotic disorder, unspecified
Substance-use -disorders	F19.96	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced persisting amnestic disorder
Substance-use -disorders	F19.97	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced persisting dementia
Substance-use -disorders	F19.980	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced anxiety disorder
Substance-use -disorders	F19.981	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced sexual dysfunction
Substance-use -disorders	F19.982	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced sleep disorder

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F19.988	ICD-10	Other psychoactive substance use, unspecified with other psychoactive substance-induced disorder
Substance-use -disorders	F19.99	ICD-10	Other psychoactive substance use, unspecified with unspecified psychoactive substance-induced disorder

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

eTable 7: U.S. State of Residence

State of residence	Full name	Number of patients	Percentage
CO	Colorado	125,267	23.0%
NY	New York	103,016	18.9%
MO	Missouri	98,450	18.1%
NC	North Carolina	75,626	13.9%
IN	Indiana	52,699	9.7%
VA	Virginia	23,848	4.4%
KY	Kentucky	18,102	3.3%
VT	Vermont	8,473	1.6%
DE	Delaware	7,313	1.3%
AK	Alaska	4,776	0.9%
MT	Montana	4,696	0.9%
AL	Alabama	3,459	0.6%
NJ	New Jersey	2,074	0.4%
IL	Illinois	1,994	0.4%
SC	South Carolina	1,193	0.2%
CT	Connecticut	1,125	0.2%
KS	Kansas	1,007	0.2%
FL	Florida	932	0.2%
NH	New Hampshire	695	0.1%
CA	California	679	0.1%
GA	Georgia	536	0.1%
PA	Pennsylvania	506	0.1%
TX	Texas	473	0.1%
MD	Maryland	451	0.1%
TN	Tennessee	448	0.1%
OH	Ohio	309	0.1%
WV	West Virginia	297	0.1%
MI	Michigan	272	0.1%
MA	Massachusetts	258	0.0%
WY	Wyoming	213	0.0%
MN	Minnesota	199	0.0%
AZ	Arizona	170	0.0%
WA	Washington	147	0.0%
NM	New Mexico	135	0.0%
WI	Wisconsin	111	0.0%
NE	Nebraska	98	0.0%
OK	Oklahoma	90	0.0%
LA	Louisiana	88	0.0%
IA	Iowa	82	0.0%
AR	Arkansas	78	0.0%

For peer review only

Online supplementary material

OR	Oregon	77	0.0%
SD	South Dakota	73	0.0%
UT	Utah	71	0.0%
DC	District of Columbia	67	0.0%
ND	North Dakota	63	0.0%
NV	Nevada	60	0.0%
MS	Mississippi	53	0.0%
RI	Rhode Island	42	0.0%
ME	Maine	42	0.0%
PR	Puerto Rico	30	0.0%
ID	Idaho	26	0.0%
HI	Hawaii	23	0.0%
AE	Armed Forces Europe	8	0.0%
AP	Armed Forces Pacific	1	0.0%

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

eTable 8: Mean CGI-S score by diagnosis

Diagnosis	Number of patients	Mean CGI-S	95% Confidence Interval	Standard deviation
Dementia & Alzheimer's disease	5,029	4.33	4.30 to 4.37	1.36
Substance-use disorders	152,790	3.61	3.60 to 3.62	1.80
Schizophrenia	14,592	4.44	4.41 to 4.46	1.44
Schizoaffective disorder	15,044	4.48	4.45 to 4.50	1.42
Bipolar disorder	69,607	4.28	4.27 to 4.29	1.45
Major depressive disorder (MDD)	129,120	4.21	4.20 to 4.22	1.42
Anxiety disorders	103,923	4.02	4.01 to 4.03	1.53
Attention-deficit hyperactivity disorder (ADHD)	53,744	3.99	3.97 to 4.00	1.42
Note: Confidence intervals were estimated using the formula mean +/- error margin where: Error margin = std/sqrt(N) * Z, Z = 1.96 for 95% confidence interval				

For peer review only

Online supplementary material

eTable 9: Mean maximum and minimum CGI-S score by diagnosis

Diagnosis	Number of patients with at least 2 CGI-S data points	Max CGI-S Mean	Min CGI-S Mean	Max CGI-S Std dev	Min CGI-S Std dev	Max CGI-S Confidence Interval	Min CGI-S Confidence Interval
Dementia & Alzheimer's disease	2,980	5.10	3.32	1.06	1.56	5.06 to5.14	3.26 to 3.37
Substance-use disorders	107,939	4.85	2.55	1.32	1.62	4.84 to 4.85	2.54 to 2.56
Schizophrenia	12,465	5.41	2.90	1.01	1.55	5.40 to 5.43	2.87 to 2.93
Schizoaffective disorder	13,590	5.53	2.88	0.95	1.55	5.51 to 5.55	2.86 to 2.91
Bipolar disorder	59,222	5.22	3.02	0.98	1.54	5.21 to 5.23	3.00 to 3.03
Major depressive disorder (MDD)	103,933	5.02	3.03	0.98	1.53	5.01 to 5.03	3.03 to 3.04
Anxiety disorders	89,710	5.04	2.80	0.97	1.58	5.03 to 5.04	2.79 to 2.81
Attention-deficit hyperactivity disorder (ADHD)	47,591	4.92	2.75	1.00	1.50	4.91 to 4.93	2.73 to 2.76

Note: Confidence intervals were estimated using the formula mean +/- error margin where:
Error margin = std/sqrt(N) * Z, Z = 1.96 for 95% confidence interval

For peer review only

Online supplementary material

eTable 10: Number of patients with at least one hospital visit by diagnosis

Diagnosis	Number of patients	Number of hospitalized patients	Percentage
Dementia & Alzheimer's disease	5029	2625	52.2%
Substance-use disorders	152790	54267	35.5%
Schizophrenia	14592	8123	55.7%
Schizoaffective disorder	15044	8528	56.7%
Bipolar disorder	69607	25322	36.4%
Major depressive disorder (MDD)	129120	40726	31.5%
Anxiety disorders	103923	22455	21.6%
Attention-deficit hyperactivity disorder (ADHD)	53744	8587	16.0%
Note: Hospitalization includes both Inpatient and Emergency services			

For peer review only

Online supplementary material

eTable 11: SAV features included in MSE NLP analysis (eTable 12)

Category	Description	SAV number
Delusions	experiencing delusions/abnormal thoughts (grandeur)	sav_1_aopt_1
Delusions	experiencing delusions/abnormal thoughts (persecution)	sav_2_aopt_2
Delusions	experiencing delusions/abnormal thoughts (religious)	sav_3_aopt_3
Delusions	experiencing hallucinations (auditory)	sav_4_aopt_4
Delusions	experiencing hallucinations (visual)	sav_5_aopt_5
Delusions	experiencing hallucinations (tactile)	sav_6_aopt_6
Delusions	History of delusions/abnormal thoughts	sav_7_aopt_7
Delusions	History of hallucinations	sav_8_aopt_8
Delusions	experiencing delusions (not specified)	sav_10_aopt_10
Delusions	experiencing hallucinations (not specified)	sav_11_aopt_11
Delusions	experiencing delusions/abnormal thoughts (paranoia)	sav_13_aopt_13
Delusions	experiencing hallucinations (olfactory)	sav_14_aopt_14
Delusions	responding to internal stimuli	sav_15_aopt_15
Delusions	experiencing delusions/abnormal thoughts (sexual)	sav_16_aopt_16
Mood	"anxious, tense"	sav_158_mood_2
Mood	"depressed, sad, despondent"	sav_159_mood_3
Mood	"irritable, angry"	sav_160_mood_4
Mood	declined	sav_162_mood_6
Mood	"elevated, manic"	sav_163_mood_7
Mood	labile	sav_164_mood_8
Mood	misc. issues	sav_165_mood_9
Mood	suicidal	sav_166_mood_10
Cognition	general issues	sav_65_attn_concentration_2
Cognition	declined	sav_67_attn_concentration_4
Cognition	issues due to MH	sav_69_attn_concentration_6
Cognition	issues due to DD	sav_70_attn_concentration_7
Cognition	varying	sav_71_attn_concentration_8
Cognition	issues with concentration	sav_81_cognition_2
Cognition	issues with fund of knowledge	sav_83_cognition_4
Cognition	issues with attention	sav_85_cognition_6
Cognition	general issues	sav_87_cognition_8
Cognition	some impairment	sav_90_exec_functioning_2
Cognition	much impairment	sav_91_exec_functioning_3
Cognition	declined since last session	sav_93_exec_functioning_5
Cognition	below age level	sav_95_exec_functioning_7
Cognition	impaired (due to DD)	sav_96_exec_functioning_8
Cognition	impaired (due to TBI)	sav_97_exec_functioning_9
Cognition	impaired (due to other MH issue)	sav_98_exec_functioning_10

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

eTable 12: Number of patients with at least one MSE feature by diagnosis

Diagnosis	MSE Category	Number of patients	Number of patients with at least one MSE feature	Percentage
Dementia & Alzheimer's disease	Delusions & hallucinations	5,029	2,147	42.7%
Dementia & Alzheimer's disease	Mood	5,029	2,518	50.1%
Dementia & Alzheimer's disease	Cognition	5,029	2,719	54.1%
Substance-use disorders	Delusions & hallucinations	152,790	37,455	24.5%
Substance-use disorders	Mood	152,790	89,616	58.7%
Substance-use disorders	Cognition	152,790	26,461	17.3%
Schizophrenia	Delusions & hallucinations	14,592	11,513	78.9%
Schizophrenia	Mood	14,592	10,753	73.7%
Schizophrenia	Cognition	14,592	6,418	44.0%
Schizoaffective disorder	Delusions & hallucinations	15,044	11,885	79.0%
Schizoaffective disorder	Mood	15,044	12,278	81.6%
Schizoaffective disorder	Cognition	15,044	6,785	45.1%
Bipolar disorder	Delusions & hallucinations	69,607	33,179	47.7%
Bipolar disorder	Mood	69,607	55,472	79.7%
Bipolar disorder	Cognition	69,607	23,373	33.6%
Major depressive disorder (MDD)	Delusions & hallucinations	129,120	42,566	33.0%
Major depressive disorder (MDD)	Mood	129,120	100,182	77.6%
Major depressive disorder (MDD)	Cognition	129,120	28,986	22.4%
Anxiety disorders	Delusions & hallucinations	103,923	32,059	30.8%
Anxiety disorders	Mood	103,923	76,320	73.4%
Anxiety disorders	Cognition	103,923	25,625	24.7%
Attention-deficit hyperactivity disorder (ADHD)	Delusions & hallucinations	53,744	11,933	22.2%
Attention-deficit hyperactivity disorder (ADHD)	Mood	53,744	31,112	57.9%
Attention-deficit hyperactivity disorder (ADHD)	Cognition	53,744	18,432	34.3%

For peer review only

Online supplementary material

eFigure 1: MindLinc EHR data de-identification and harmonization pipeline

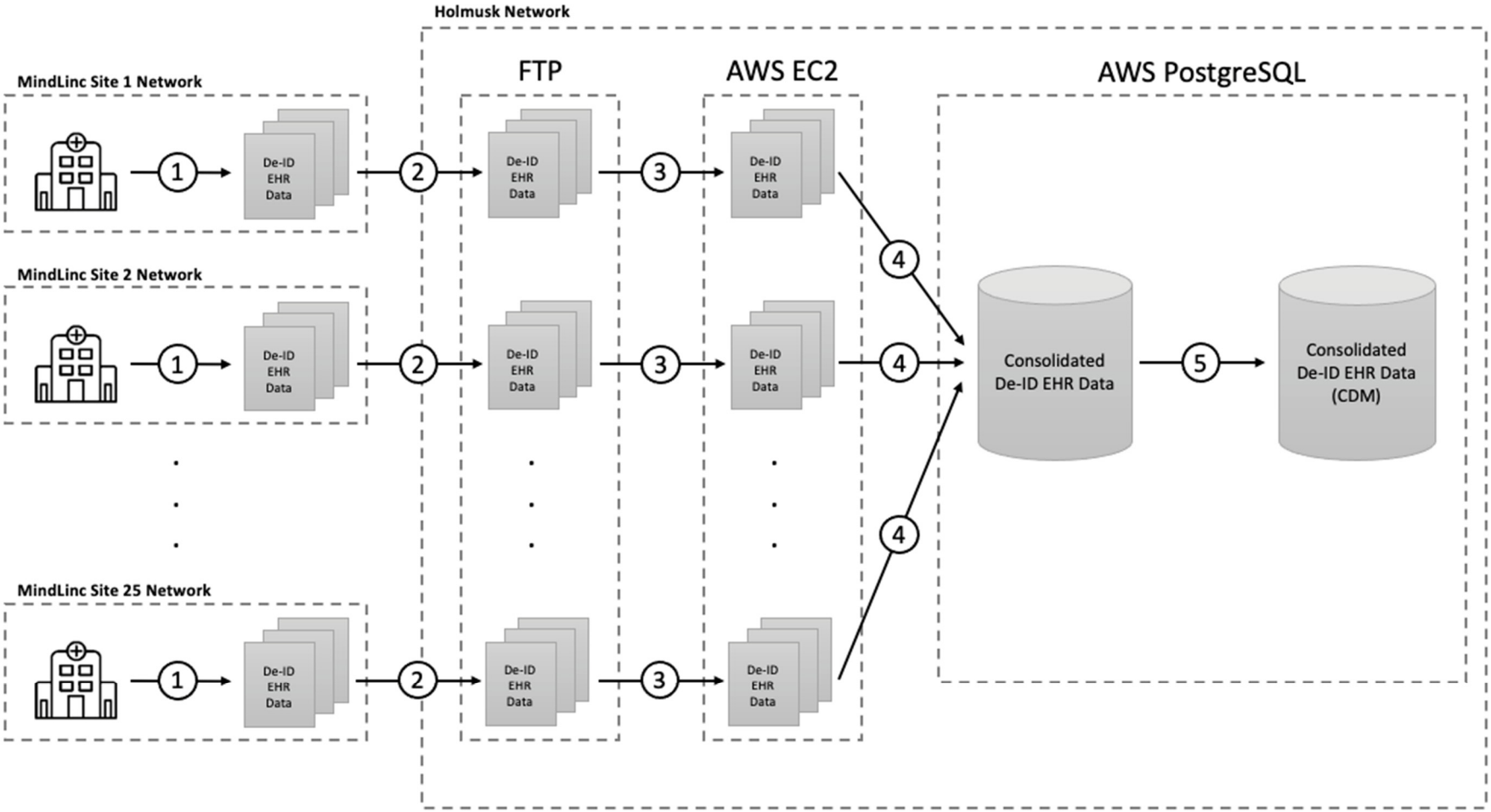


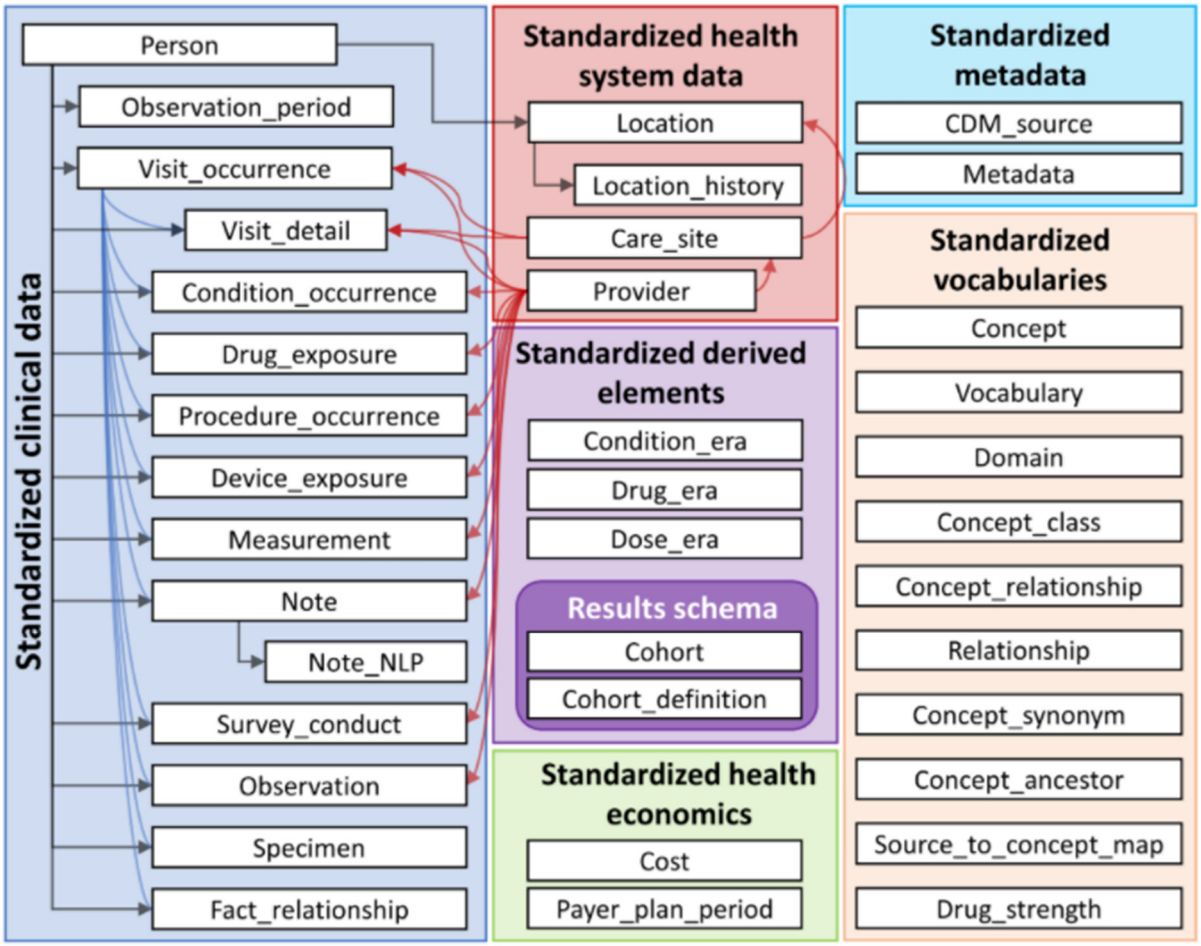
Figure: MindLinc Data Flow Diagram

Online supplementary material

eFigure 2: OMOP CDM schematic

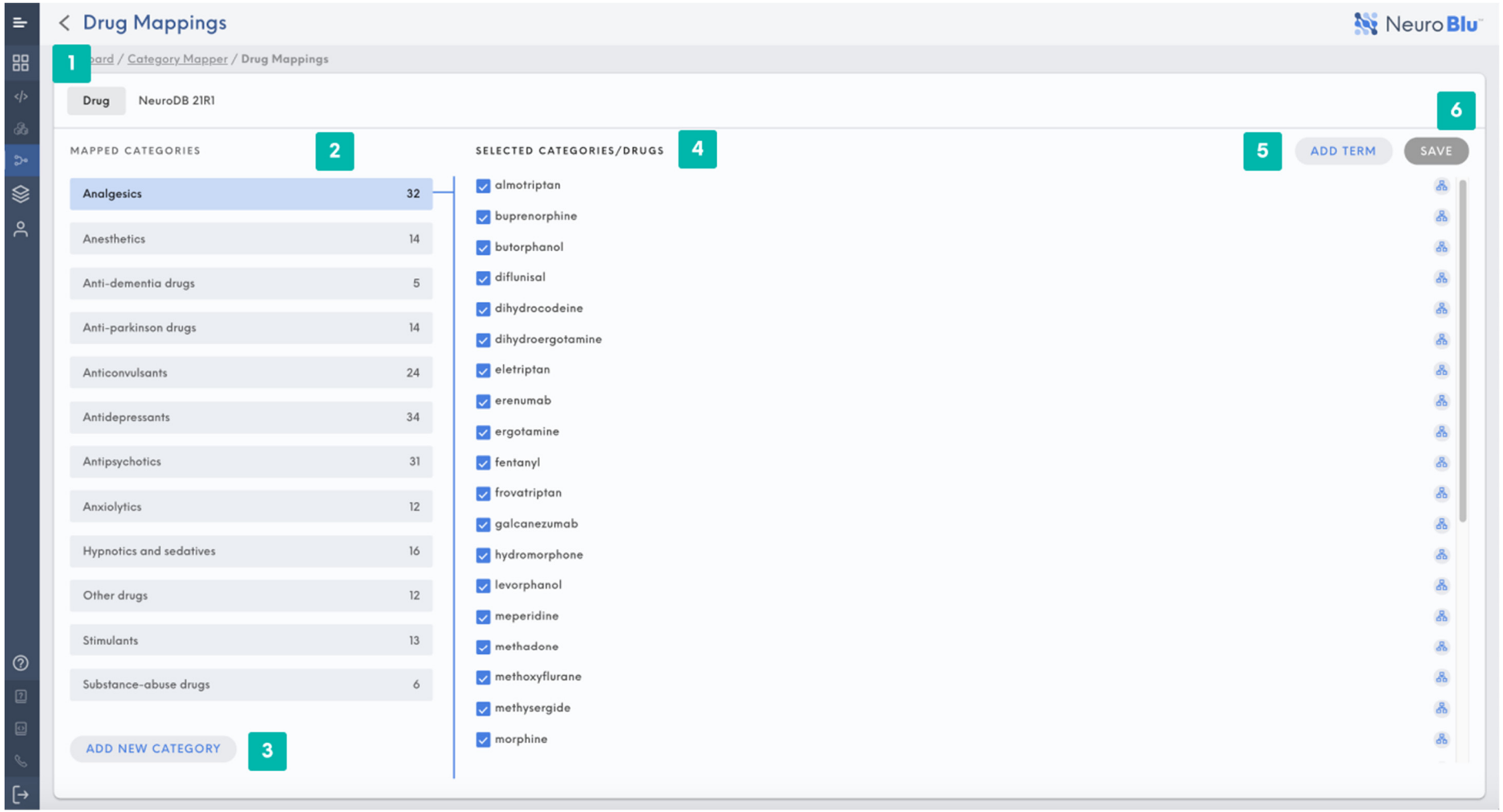
Source: https://ohdsi.github.io/TheBookOfOhdsi/CommonDataModel.html

Attribution: Creative Commons Zero v1.0 Universal license (No copyright)



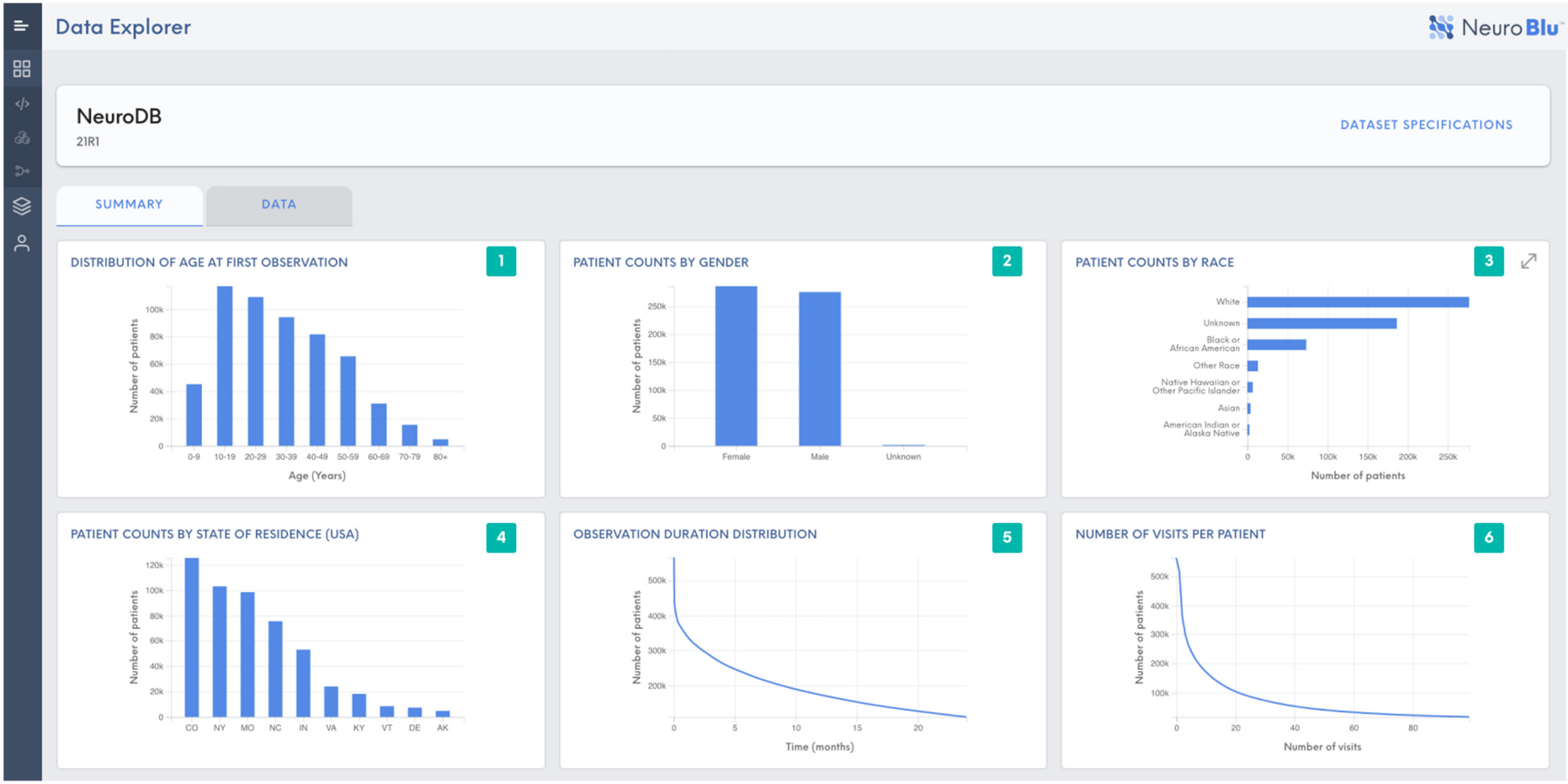
Online supplementary material

eFigure 3: NeuroBlu Category Mapper illustrating the classification of individual drugs into different types of medication



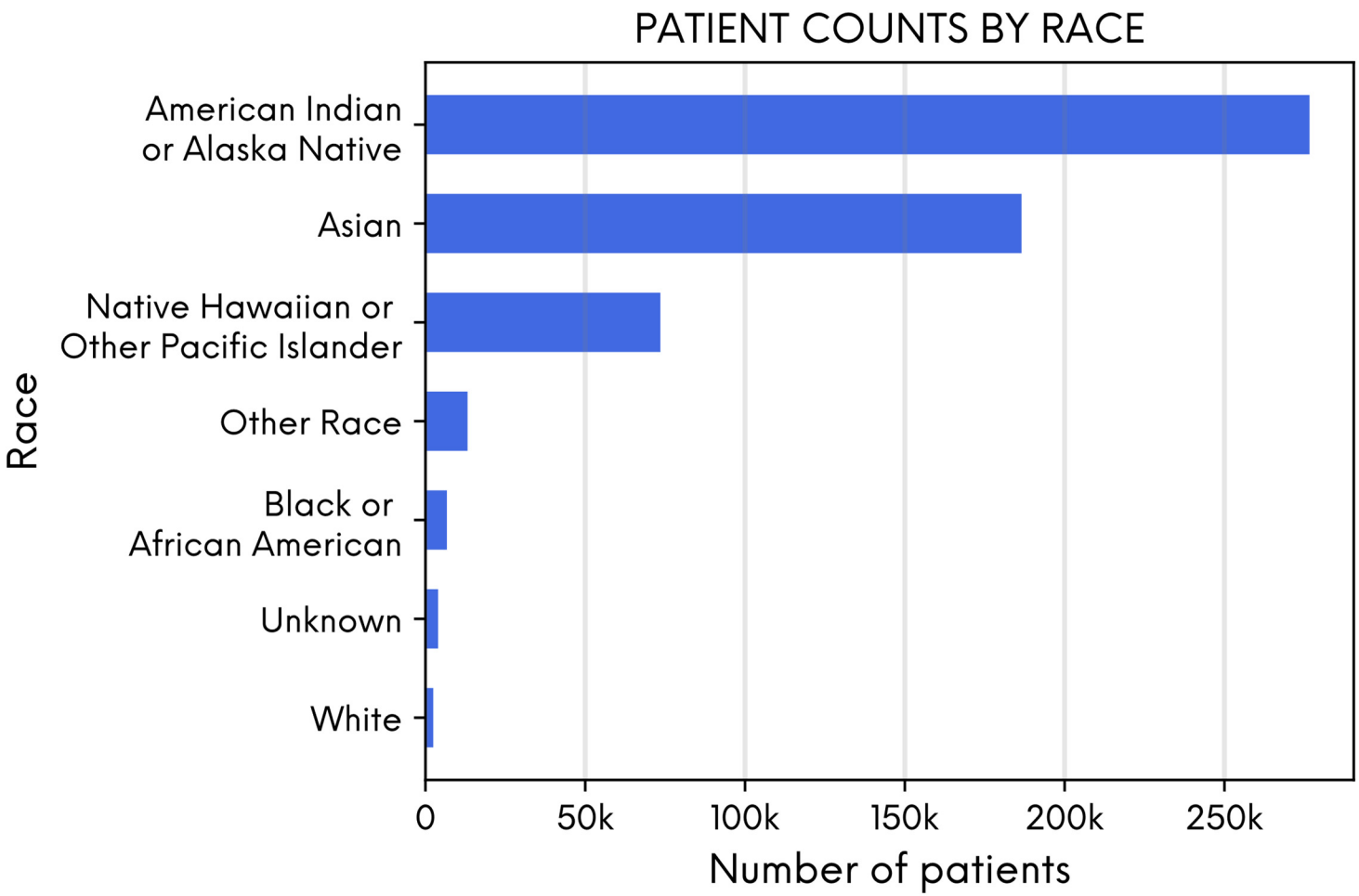
Online supplementary material

eFigure 4: NeuroBlu Data Explorer which provides a summary of the dataset and allows the top 50 rows of each table to be viewed

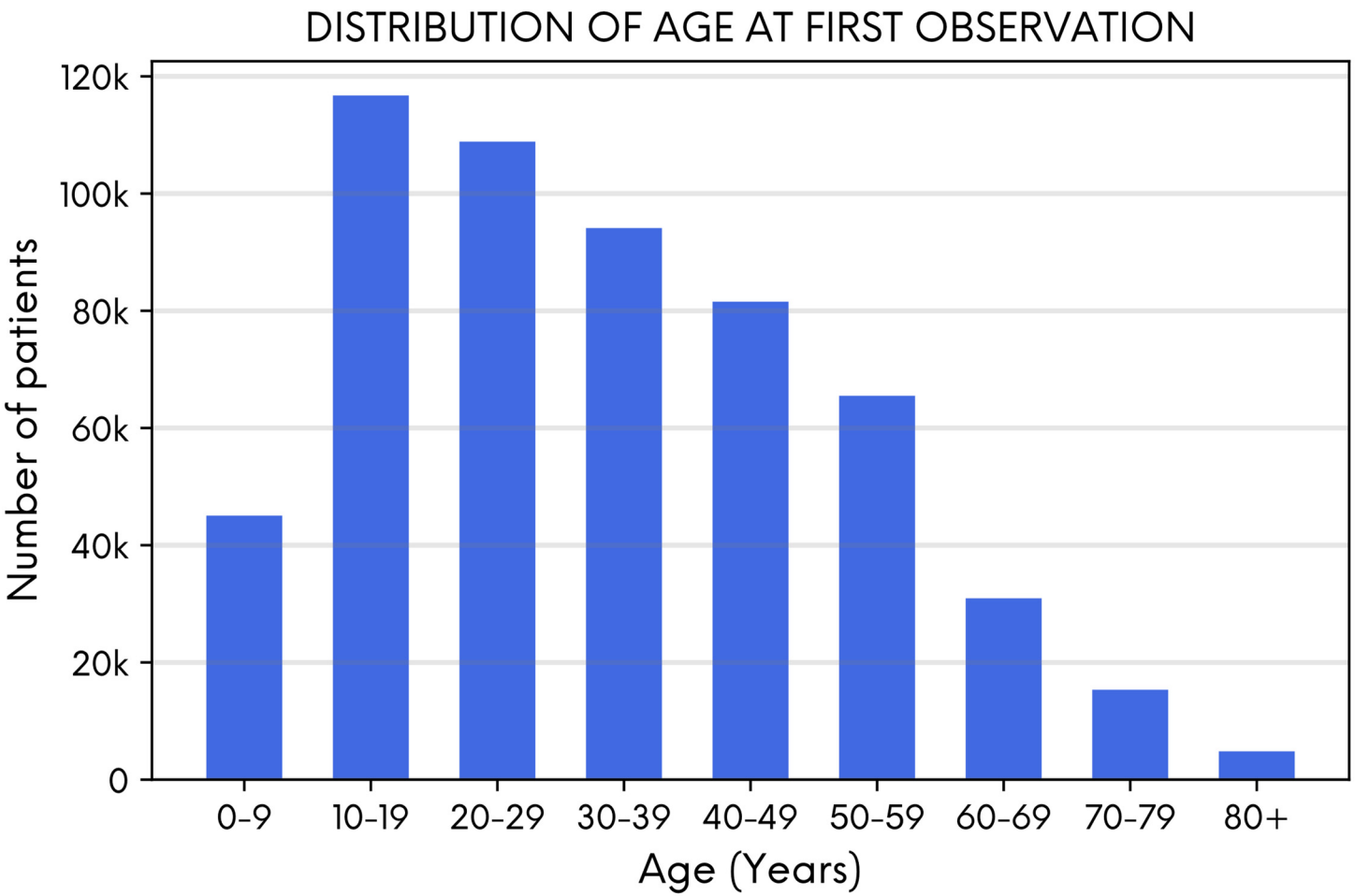


Online supplementary material

eFigure 5: NeuroBlu dataset race distribution

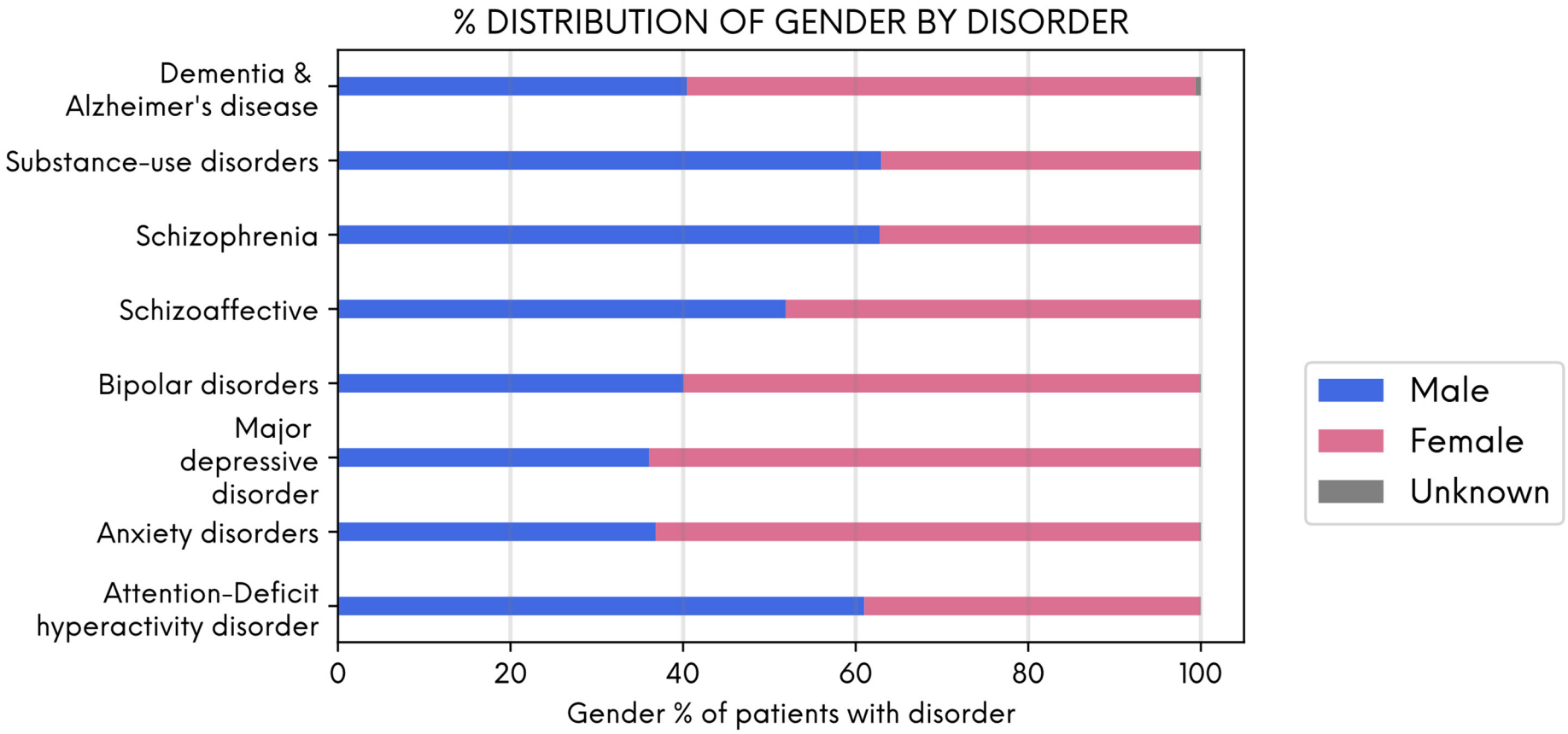


eFigure 6: NeuroBlu dataset age distribution

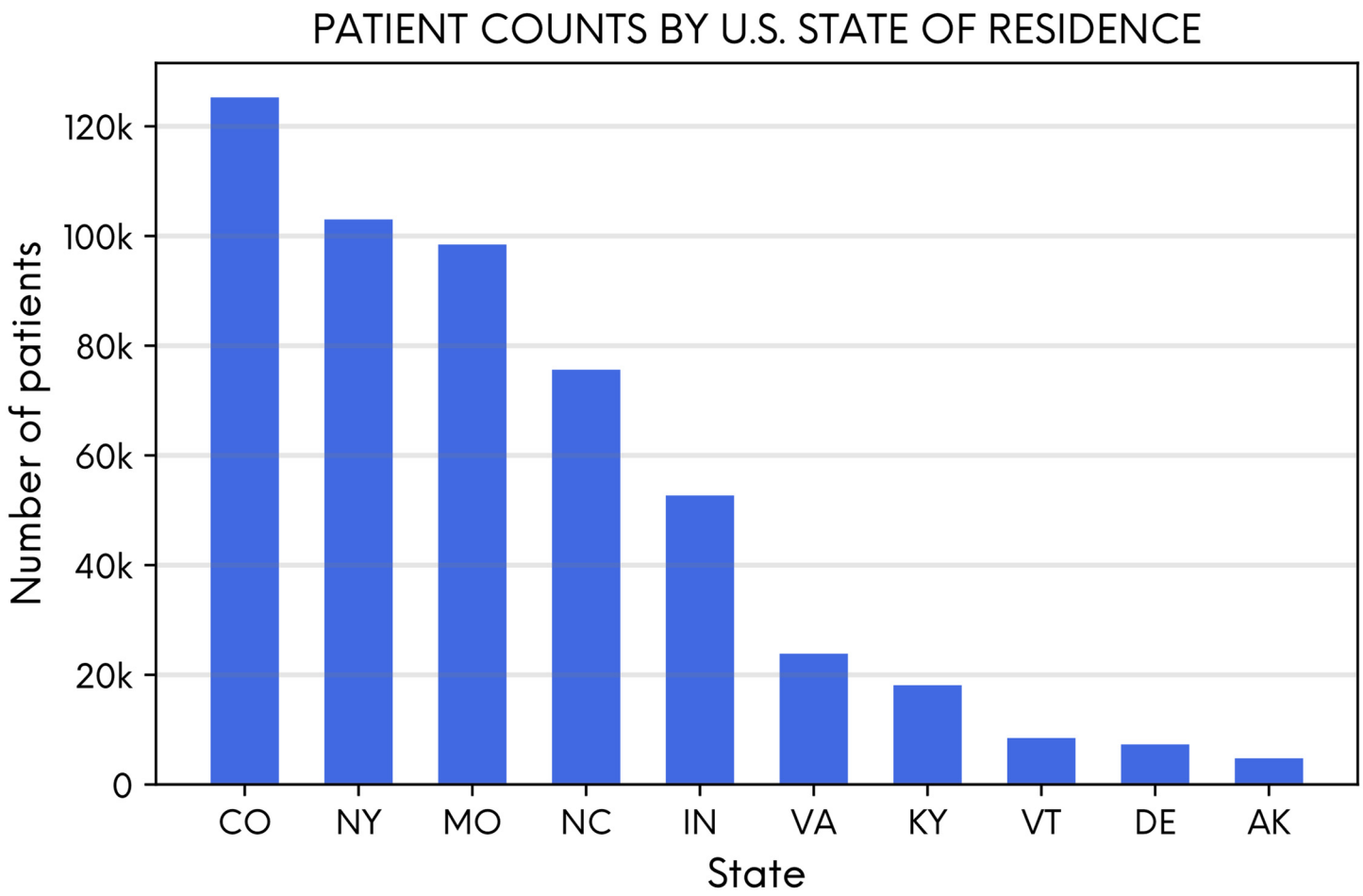


Online supplementary material

eFigure 7: Gender distribution by diagnosis

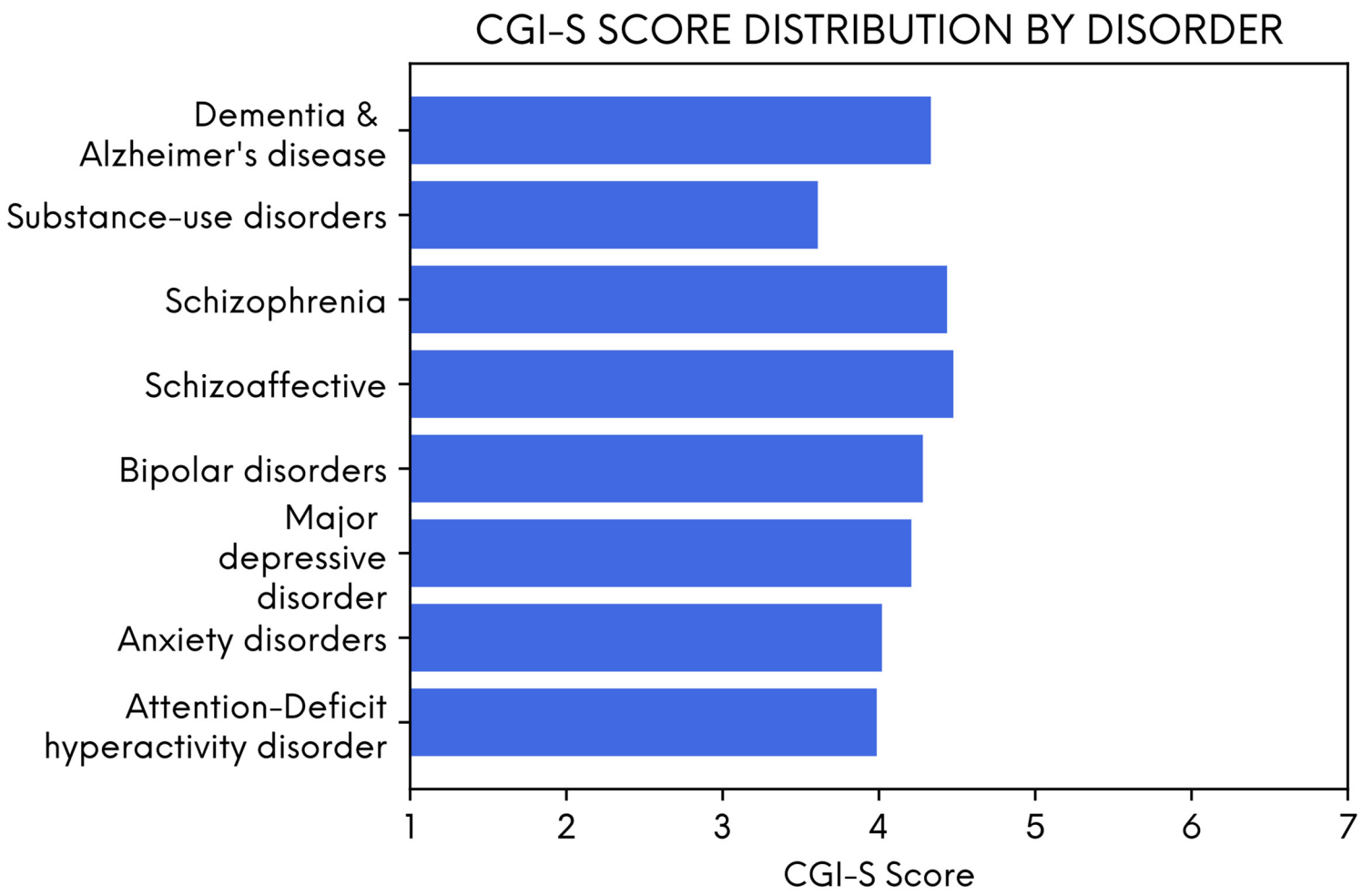


eFigure 8: U.S. State of Residence



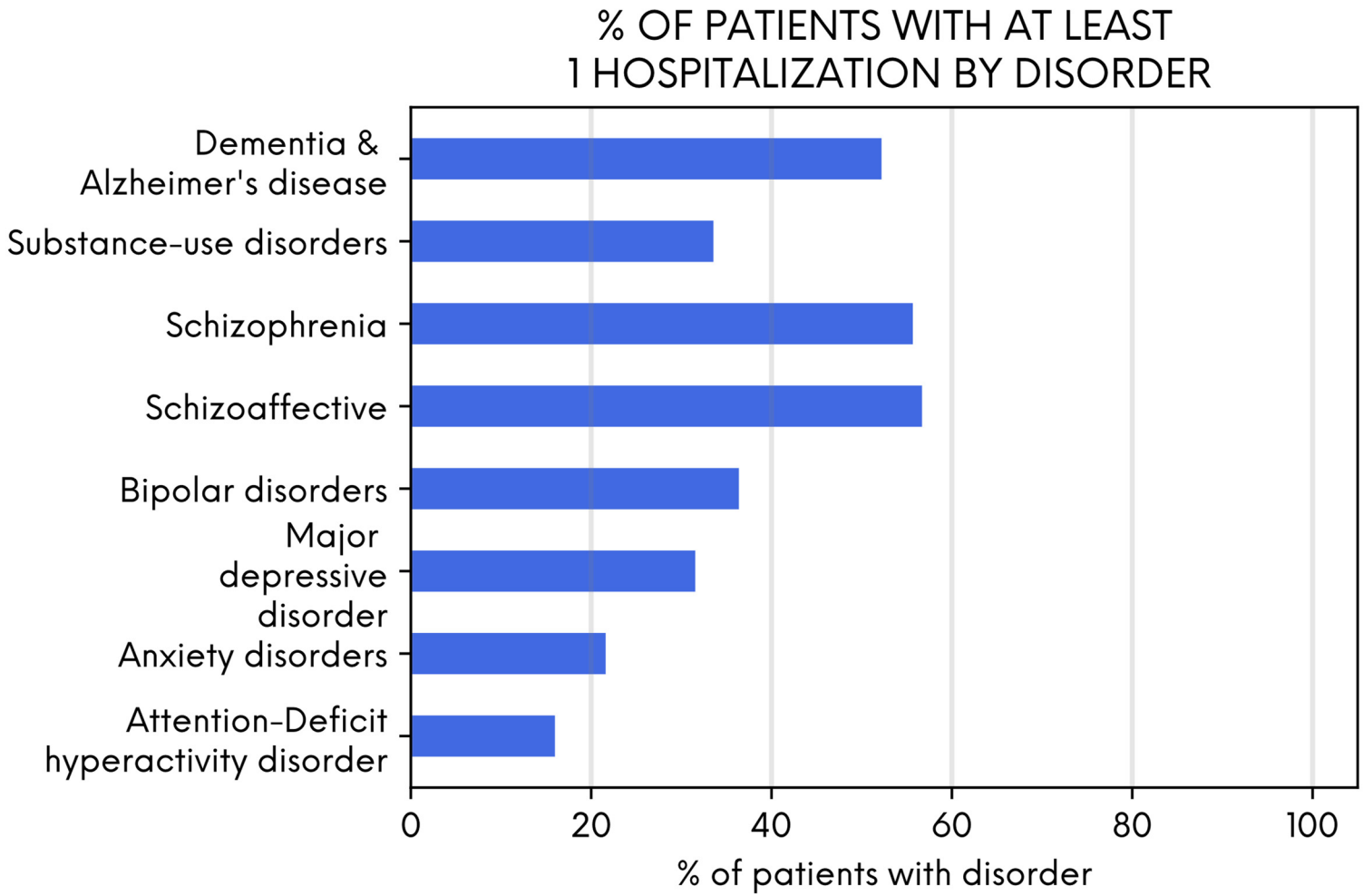
Online supplementary material

eFigure 9: Mean CGI-S score by diagnosis



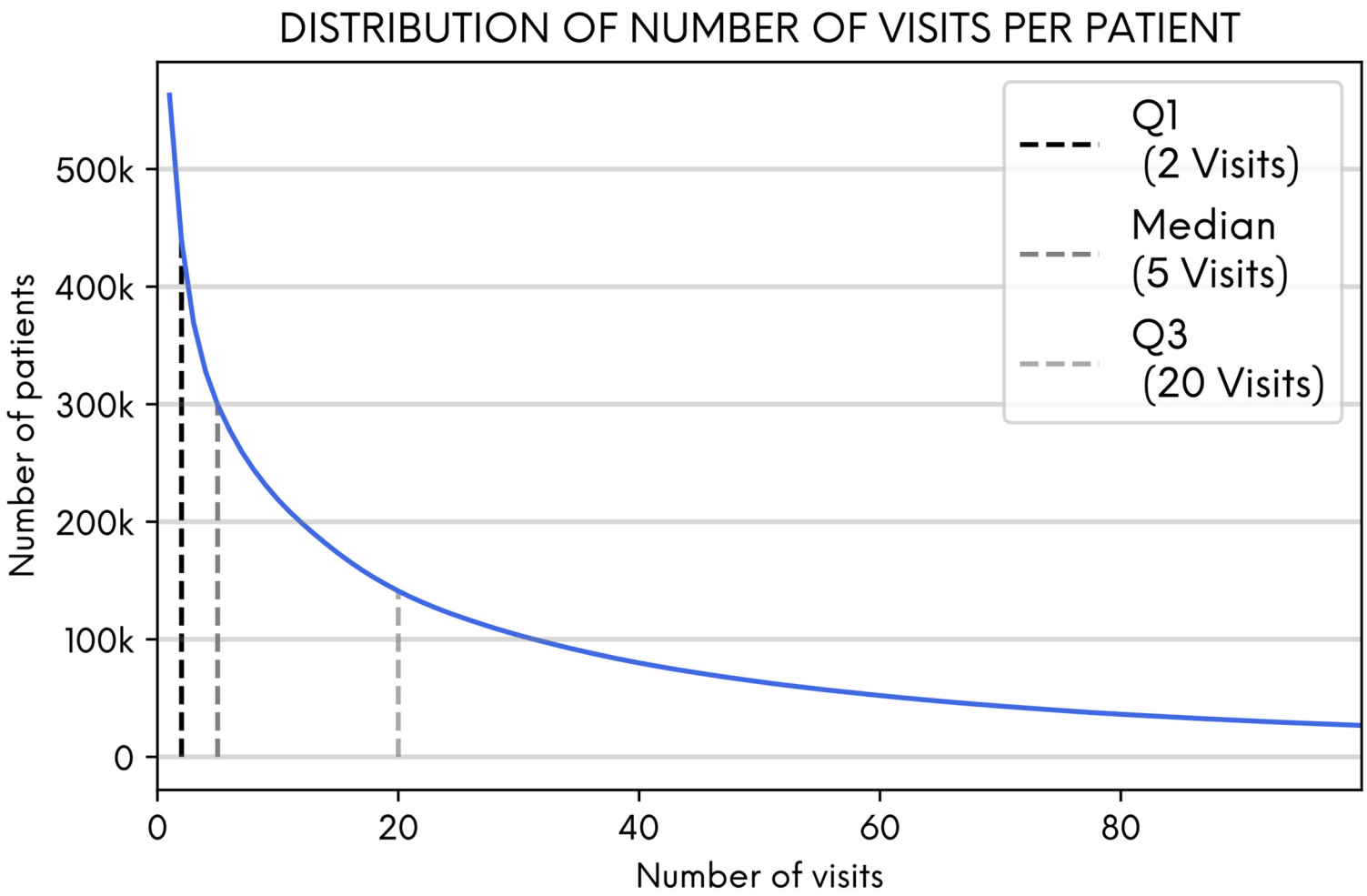
Online supplementary material

eFigure 10: Percentage of patients with at least one hospital visit by diagnosis

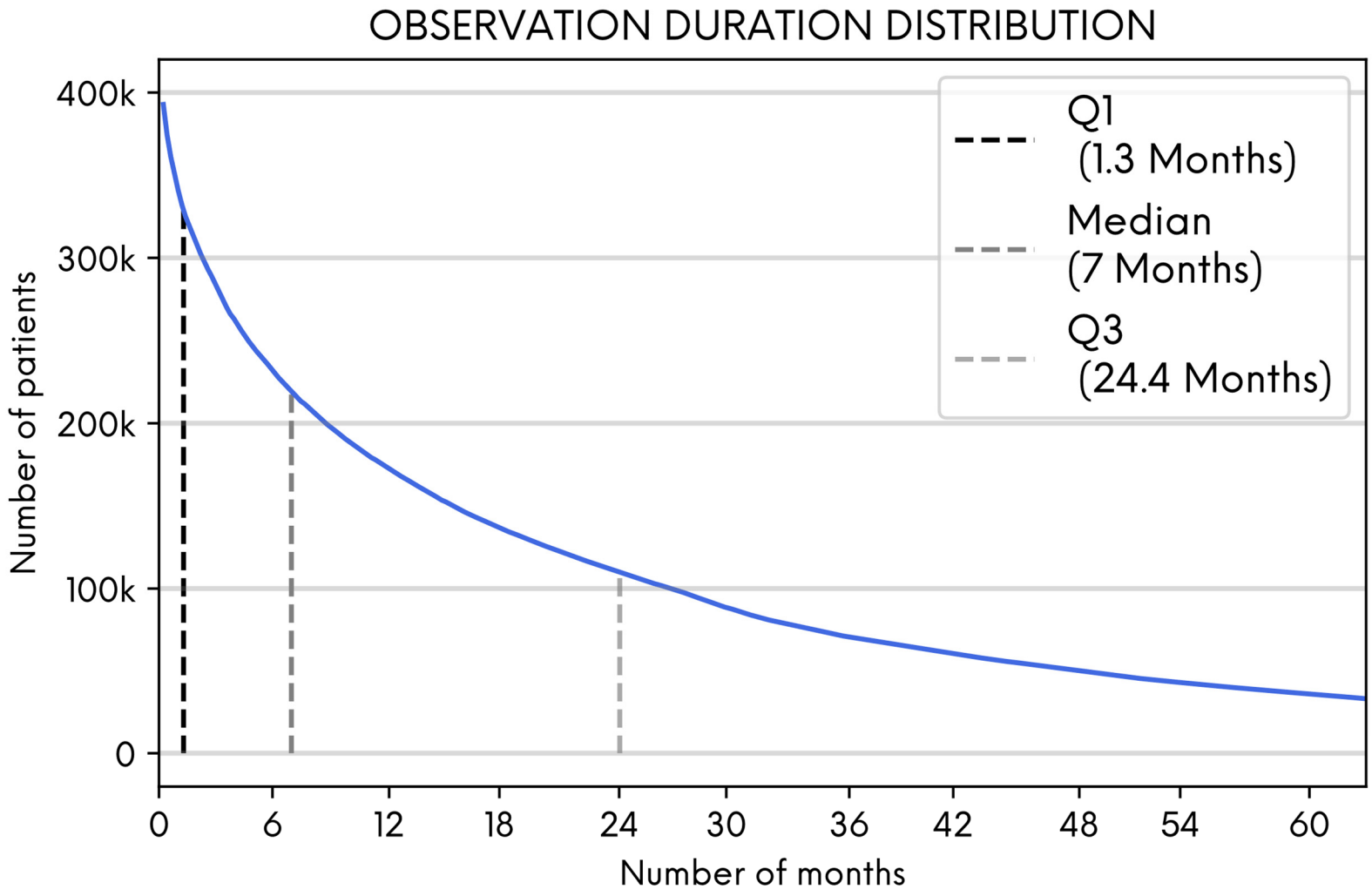


Online supplementary material

eFigure 11: Distribution of number of visits per patient



eFigure 12: Distribution of duration of follow-up



BMJ Open

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare analytics with real-world data

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-057227.R1
Article Type:	Cohort profile
Date Submitted by the Author:	18-Jan-2022
Complete List of Authors:	Patel, Rashmi; Holmusk USA Inc; King's College London Institute of Psychiatry Psychology and Neuroscience, Department of Psychosis Studies Wee, Soon Nan; Holmusk USA Inc Ramaswamy, Rajagopalan; Holmusk USA Inc Thadani, Simran; Holmusk USA Inc Tandi, Jesisca; Holmusk USA Inc Garg, Ruchir; Holmusk USA Inc Calvanese, Nathan; Holmusk USA Inc Valko, Matthew; Holmusk USA Inc Rush, A.; Curbstone Consultant LLC Rentería, Miguel; Holmusk USA Inc Sarkar, Joydeep; Holmusk USA Inc Kollins, Scott; Holmusk USA Inc; Duke University School of Medicine
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Health informatics
Keywords:	PSYCHIATRY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, EPIDEMIOLOGY

SCHOLARONE™
Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

For peer review only

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment 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] and Scott H. Kollins[1,4]

1. Holmusk, 54 Thompson Street, New York, NY 10012, USA
2. King's College London (Institute of Psychiatry, Psychology and Neuroscience), London, UK
3. Curbstone Consultant LLC, Santa Fe, NM, USA
4. Duke University School of Medicine (Department of Psychiatry and Behavioral Sciences), Durham, NC, USA

Correspondence to:

Rashmi Patel

Holmusk, 54 Thompson Street, New York, NY 10012, USA

E-mail: rashmi.patel@holmusk.com

Manuscript word count (excluding references): 5,656

Abstract word count: 298

For peer review only

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare real-world data analytics

Abstract

Purpose:

NeuroBlu is a real-world data (RWD) repository that contains de-identified electronic health record (EHR) data from U.S. 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, ICD-9/ICD-10 diagnosis, prescribed medications, family history of mental disorders, Clinical Global Impression (CGI-S/CGI-I), and Global Assessment of Functioning (GAF). To further enhance the dataset, 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 dataset supports the generation of real-world evidence (RWE) in mental health.

Participants:

As of 31st July 2021, 562,940 individuals (48.9% male) were present in the dataset with a mean age of 33.4 years (standard deviation: 18.4 years). The most frequently recorded diagnoses were substance use disorders (152,790 patients), major depressive disorder (129,120 patients) and anxiety disorders (103,923 patients). The median duration of follow-up was 7 months (interquartile range: 1.3 to 24.4 months).

Findings to date:

The dataset has supported epidemiological studies demonstrating increased risk of psychiatric hospitalization and reduced antidepressant treatment effectiveness among people with comorbid substance use disorders. It has also been used to develop data visualization 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 dataset will be further analyzed 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 health care services.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare real-world data analytics

Article Summary

Strengths and limitations

- The NeuroBlu dataset benefits from a large sample size of de-identified electronic health record (EHR) data from over 560,000 people who have received mental healthcare over a period of 21 years.
- The dataset is built on a robust de-identification pipeline and encryption framework that enables a wide range of users to safely analyze data through a trusted research environment using a graphical user interface or advanced analytic software (R and Python).
- Structured data on clinical severity (CGI-S) are recorded for over 80% of patients in the dataset. Natural language processing (NLP) enables access to rich clinical data from unstructured free text data as part of the Mental State Examination (MSE) and social factors that are not typically available in purely structured clinical datasets such as claims or RCT data.
- As the dataset 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.
- 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.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 U.S. 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 expectancy[3] and the lifetime costs associated with SMI are estimated to be $1.85 million 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 datasets, known as real world data (RWD), has improved the understanding of a range of medical disorders at a population level[7] and contributed to the development[8] and evaluation[9] of novel therapeutics. RWD may be analyzed to generate real-world evidence (RWE) to quantify the safety and effectiveness of therapeutic interventions. The generation of RWE from the analysis of largescale datasets that are representative of healthcare service delivery in routine settings provides a complementary source of data – as compared to more controlled clinical trials – that 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 characterize health disorders and to develop and deliver more effective treatments.[12] EHR datasets have previously supported RWE generation in critical care[13] and infrastructure based on common data models[14] has been developed to enable researchers to securely analyze 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-standardized 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 analyzed. Relatedly, there is little standardization 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 de-identified mental health EHR datasets have been curated to support largescale population research,[19–21] provide clinical phenotype data for genomics research,[22] 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

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

unstructured free text in EHRs and enable statistical analysis of detailed clinical data such as presenting symptoms,[27] environmental factors,[28] 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 datasets are only accessible by clinical researchers based in specialized academic health science centers.[31] Enabling wider access, while maintaining robust data privacy and security standards, could substantially increase the impact of de-identified EHR datasets 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 de-identified EHR data from U.S. mental healthcare providers, including state-of-the-art NLP software to characterize mental state information from semi-structured 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.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Cohort Description

Participants and setting

The NeuroBlu platform enables users to analyze de-identified data from over 560,000 patients receiving care from 25 U.S. mental healthcare providers that operate the MindLinc EHR system. Supplementary eTable 1 provides details of the locations, types of mental healthcare service and numbers of patients for each provider. In summary, the platform includes de-identified EHR data from individuals receiving mental healthcare from outpatient, inpatient, telemedicine, and residential care facilities in 12 U.S. 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 semi-structured free text fields to document the mental state examination (MSE) and treatment plan.[33] A subset of de-identified 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 de-identified at source before being transformed and normalized in a cloud-based U.S. Amazon Web Services (AWS) data warehouse and mapped to a common data model to create a harmonized dataset including data from all participating healthcare providers (eFigure 1). NeuroBlu users analyze the harmonized dataset 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, de-identified EHR data may be analyzed by anyone with access to the trusted research environment without requiring any movement or disclosure of the underlying data.

De-identification procedure

MindLinc EHR data are de-identified at source (i.e. within the computing infrastructure of each participating healthcare provider) using the Safe Harbor method (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 dataset. 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, URLs, social security numbers, 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 randomizing identifiers, recording age in years, and collecting the first three digits of zip codes with a population greater than 20,000).

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Transfer of de-identified EHR data from healthcare provider to secure AWS cloud

The de-identified EHR data are compressed and transferred from the source healthcare provider through a virtual private network (VPN) to a secure file transfer protocol (FTP) site (eFigure 1, step 2). The data are then transferred to a secure Amazon Elastic Compute Cloud (AWS EC2) instance (eFigure 1, step 3) where they are decompressed and assembled for loading into a PostgreSQL database (eFigure 1, step 4).

Given the relational nature of the MindLinc databases, the files produced by the EHR data de-identification 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 de-identification process is irreversible, preventing re-identification 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 (eFigure 1, step 4) and is further described in section 1 of the Supplementary 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 (e.g. appearance and behavior) or may be elicited through direct questioning or the presentation of cognitive tasks. The MSE provides a rich source of clinical phenotype data which characterizes the nature of presenting symptoms which could be associated with varying clinical outcomes and response to treatment.

The MindLinc EHR includes a semi-structured "status assessment" field in which clinicians can document features associated with a patient's MSE. The status assessment field allows clinicians to choose pre-defined features from a list of options and/or to document findings as unstructured free text. However, pre-defined 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 pre-defined features in MindLinc is 0% whereas the median percentage with unstructured free text is 67%.

NLP is a text mining technique that enables 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 involves a series of processes to develop algorithms that can identify clinically meaningful concepts for secondary analysis. These processes include: (i) Data assembly: identifying a collection of relevant documents (the corpus), (ii) 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, (iii) Preprocessing: preparation of the corpus for NLP model development including stop word removal, stemming, lemmatization and parts of speech tagging, (iv) Featurization: classifying text within the corpus into different features (e.g. parts of speech, word vectors or embeddings, sentiment or temporal features) and (v) Analysis: development of NLP algorithm using a rules-based or machine learning approach using the training set. The resulting models are evaluated on the previously annotated reference set and tuned to maximize accuracy as measured through precision (positive predictive value), recall (sensitivity) and F1 measure (harmonic mean of precision and recall).[37] If a sufficiently accurate models are developed, they can be applied to the entire corpus to generate structured data on clinically meaningful features of

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

interest[27] 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 de-identified EHR dataset to create a table of NLP-derived MSE measures which may be joined and analyzed with other structured clinical and sociodemographic data in the NeuroBlu dataset. 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 below in the section on *Unstructured data*.

Data variables

Full details of available data variables and their structure within the NeuroBlu MindLinc dataset are provided in section 2 of the Supplementary 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 dataset includes structured sociodemographic data on gender, year of birth, race, ethnicity, U.S. 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 dataset that include data on contacts with mental health services, emergency room visits, hospital admissions, 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 dataset. Structured questionnaire data are available for the following numbers of patients within the dataset:

(i) CGI-S: 471,256 patients (83.7%)
(ii) CGI-I: 361,819 patients (64.3%)
(iii) GAF: 310,895 patients (55.3%)
(iv) MADRS: 1,126 patients (0.2%)

The dataset 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 Supplementary Material.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 Supplementary Material. Information on NLP model development and precision statistics 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 & MSE data, only the data derived using NLP for 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; Judgment; Language; Level of consciousness; Memory; Mood; Orientation; Psychomotor; Reasoning; Sensorium; Sleep; Speech; Suicidal; 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 (e.g. 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 analyzed 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 to 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 target 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: (i) Cohort Builder, (ii) Category Mapper, (iii) Data Explorer and (iv) R/Python Code engine. These elements allow users to define a population cohort based on specified inclusion/exclusion criteria, define key variables to examine

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 dataset 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 hospitalized) 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 (eFigure 3). This enables users to pre-define 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 dataset is provided in the Data Explorer with accompanying Dataset Specifications that describe the data structure (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 dataset 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 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: (i) *NeuroBlu*: utility functions which simplify data assembly processes using prebuilt SQL queries, (ii) *SurvBlu*: a framework for conducting time-to-event analyses, and (iii) *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 visualizing data.

Data privacy and ethics

Under HIPAA regulations, U.S. healthcare providers record and store personal healthcare information for the purposes of providing treatment, billing data, to support healthcare service

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

operations. De-identified data may be used for secondary analyses. The NeuroBlu platform has received a waiver of authorization for analysis of de-identified healthcare data from the WCG Institutional Review Board (Ref: WCG-IRB 1-1470336-1).

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 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 is encrypted with AES256.

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

Participant characteristics

Descriptive statistics for the NeuroBlu dataset were obtained on 31st July 2021. A total of 562,940 individuals (48.9% male) were present in the dataset with a mean age of 33.4 years (standard deviation: 18.4 years). The majority of the population was White (49.1%). 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 dataset, working age adults (aged 20 to 59 years) account for 62.2% and adults aged 60 years or above account for 9.1% (eTable 3/eFigure 6).

eTable 4 and Figure 3 illustrate the prevalence of mental disorders in the dataset. 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 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 & Alzheimer's disease, bipolar disorder, major depressive disorder and anxiety disorders among female patients. The constituent diagnostic codes are provided in eTable 6.

eTable 7 and eFigure 8 provide a breakdown by U.S. State of Residence. The most frequent U.S. 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 eTable 8 and eFigure 9. The greatest mean CGI-S scores were found in people with schizoaffective disorder (4.48, standard deviation: 1.42) and schizophrenia (4.44, std dev: 1.44) with the lowest mean score recorded in people with substance-use disorders (3.61, std dev: 1.80). A further analysis was conducted among patients with at least 2 recorded CGI-S scores to estimate the mean maximum and mean minimum score by diagnosis (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

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

and psychotic disorders which are typically characterized by periods of remission in between episodes.

The number of patients with at least one hospital visit (inpatient or emergency department) by diagnosis (eTable 10/eFigure 10) 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 hospitalization compared to 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 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 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.

eFigure 11 provides a distribution of the total number of clinical encounters (visits) per patient. Patients had a median of 5 visits recorded (interquartile range: 2 to 20 visits). eFigure 12 provides a distribution of the total follow-up duration available for each patient in the dataset. The median follow-up duration was 7 months (IQR: 1.3 to 24.4 months).

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

(i) Gender: 1,572 patients (0.23%)
(ii) Race: 186,517 patients (33.1%)
(iii) Psychiatric diagnosis: 66,064 patients (11.7%)
(iv) State: 21,919 patients (3.9%)
(v) NLP-derived MSE data: 132,690 patients (23.6%)

Patient and Public Involvement

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

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Findings to date

The NeuroBlu platform and MindLinc EHR dataset has been used to generate real world evidence through largescale mental health epidemiology studies as well as to develop automated NLP, data visualization and predictive analytic tools based on real world EHR data.

Epidemiology of mental disorders

Data from people with substance use disorders has been analyzed to investigate its impact among minority ethnic people,[49] and demonstrated increased rates of psychiatric hospitalization among children and adolescents[50] and increased risk of psychiatric comorbidities and hospitalization 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 has been analyzed to demonstrate that antidepressant prescribing reduced in response to U.S. 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 analyzed to generated real world evidence, the dataset has been used to develop novel data visualization and predictive analytic methods to support the analysis of mental health real world data including a data visualization 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 visualization 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 dataset, as described previously.[33,60]

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Strengths and limitations

The NeuroBlu dataset contains a rich source of de-identified EHR data that supports the generation of real-world evidence to improve understanding of mental disorders and factors affecting clinical outcomes. NeuroBlu includes data on over 560,000 people attending 25 U.S. mental healthcare providers over a period of 21 years. The large sample size and long duration of follow-up enables 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 randomized controlled trials (RCTs) or prospective observational studies (for example, comparative effectiveness studies on multiple medications or assessing the impact of substance use disorders on clinical outcomes).

NeuroBlu is built upon a robust de-identification pipeline that ensures no personal identifiable data are available to users who analyze the data. A strong data security framework that employs internationally validated encryption standards protects the anonymized dataset from being accessed outside the trusted research environment of NeuroBlu. These safeguards ensure the security of the dataset while still enabling users to easily analyze 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 visualize 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 analyze the data while ensuring the data remain secure within the trusted research environment. These features allow users from a wide range of backgrounds to analyze the data without requiring expertise in statistical software while at the same time enabling expert data scientists to perform advanced analytical procedures on the dataset using R or Python.

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

The dataset 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 dataset 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 analyzing EHR datasets 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 dataset is that clinical data may not be comprehensively completed across all domains for all patients. Clinicians typically

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

record data which 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 dataset 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 dataset 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 dataset 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 (e.g. acute physical healthcare providers or diagnoses and medications recorded in community primary care settings) are not available within the NeuroBlu dataset.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Future plans

As patients continue to receive care from healthcare providers using the MindLinc EHR, the NeuroBlu dataset will grow over time increasing in sample size and follow-up duration. We plan to analyze these data to generate real-world evidence 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, 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 personalizing 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 enables the MindLinc dataset to be analyzed by a wide range of users from varied professional disciplines in any geographical location. We actively collaborate with a 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 organizations who wish to analyze the dataset 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 de-identification pipeline and trusted research environment to support real world evidence generation from their own datasets. 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.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare real-world data analytics

Figure captions

Figure 1: NeuroBlu Cohort Builder illustrating the population characteristics of female patients diagnosed with Post Traumatic Stress Disorder who have received sertraline

Figure 2: NeuroBlu R Code Engine which 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

Figure 3: Prevalence of mental disorders in the NeuroBlu dataset

Figure 4: Mean maximum and mean minimum CGI-S score by diagnosis

Figure 5: Percentage of patients with at least one NLP-derived MSE feature by diagnosis

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Further details

Competing interests: All authors have completed the ICMJE uniform disclosure form and declare: 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.

Author contributions: 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, MR, JS and SK) contributed to cohort profile design, manuscript preparation and approved the final version.

Ethics approval: The NeuroBlu platform has received a waiver of authorization for analysis of de-identified healthcare data from the WCG Institutional Review Board (Ref: WCG-IRB 1-1470336-1).

Source of funding: The NeuroBlu dataset and MindLinc EHR system are owned and funded in full by Holmusk.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

References

1 Rehm J, Shield KD. Global Burden of Disease and the Impact of Mental and Addictive Disorders. *Curr Psychiatry Rep* 2019;**21**:10. doi:10.1007/s11920-019-0997-0

2 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. https://www.samhsa.gov/data/sites/default/files/reports/rpt29393/2019NSDUHFFRPDFWHTML/2019NSDUHFFR1PDFW090120.pdf

3 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.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1563985/

4 Seabury SA, Axeen 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. doi:10.1377/hlthaff.2018.05246

5 Degenhardt L, Charlson F, Ferrari A, *et al.* 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. *The Lancet Psychiatry* 2018;**5**:987–1012. doi:10.1016/S2215-0366(18)30337-7

6 Mental Health America. 2020 Access to Care Data. Ment. Heal. Am. 2020.https://www.mhanational.org/issues/2020/mental-health-america-access-care-data#four

7 Maissenhaelter BE, Woolmore AL, Schlag PM. Real-world evidence research based on big data: Motivation—challenges—success factors. *Onkologe* 2018;**24**:91–8. doi:10.1007/s00761-018-0358-3

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. doi:10.1016/j.drudis.2017.12.002

9 Corrigan-Curay J, Sacks L, Woodcock J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. *JAMA - J Am Med Assoc* 2018;**320**:867–8. doi:10.1001/jama.2018.10136

10 Klonoff DC. The New FDA Real-World Evidence Program to Support Development of Drugs and Biologics. *J Diabetes Sci Technol* 2020;**14**:345–9. doi:10.1177/1932296819832661

11 Roberts MH, Ferguson GT. Real-World Evidence: Bridging Gaps in Evidence to Guide Payer Decisions. *PharmacoEconomics - Open* 2021;**5**:3–11. doi:10.1007/s41669-020-00221-y

12 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–e1912869. doi:10.1001/jamanetworkopen.2019.12869

13 Johnson AEW, Pollard TJ, Shen L, *et al.* MIMIC-III, a freely accessible critical care database. *Sci Data* 2016;**3**:160035. doi:10.1038/sdata.2016.35

14 OMOP Common Data Model – OHDSI. 2022.https://www.ohdsi.org/data-standardization/the-common-data-model/

15 NORC Data Enclave | NORC.org. norc. 2022.https://www.norc.org/Research/Capabilities/Pages/data-enclave.aspx

16 Administrative Data Research Facility (ADRF) – Coleridge Initiative. 2022.https://coleridgeinitiative.org/adrf/

17 Stewart R, Davis K. 'Big data' in mental health research: current status and emerging possibilities. *Soc Psychiatry Psychiatr Epidemiol* 2016;**51**:1055–72. doi:10.1007/s00127-016-1266-8

18 Peterson BS. Editorial: Biomarkers in precision medicine for mental illnesses. *J Child Psychol Psychiatry Allied Discip* 2020;**61**:1279–81. doi:10.1111/jcpp.13357

19 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**. doi:10.1136/bmjopen-2015-008721

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare real-world data analytics

20 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. doi:10.1371/journal.pone.0190703

21 Cardinal RN. Clinical records anonymisation and text extraction (CRATE): An open-source software system. *BMC Med Inform Decis Mak* 2017;**17**:50. doi:10.1186/s12911-017-0437-1

22 Smoller JW. The use of electronic health records for psychiatric phenotyping and genomics. *Am J Med Genet Part B Neuropsychiatr Genet* 2018;**177**:601–12. doi:10.1002/ajmg.b.32548

23 Holderness E, Miller N, Cawkwell P, *et al.* Analysis of risk factor domains in psychosis patient health records. *J Biomed Semantics* 2019;**10**:19. doi:10.1186/s13326-019-0210-8

24 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* Published Online First: 2020. doi:https://doi.org/10.1016/j.schres.2020.09.016

25 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. doi:10.1016/j.schres.2020.05.007

26 Raket LL, Jaskolowski J, Kinon BJ, *et al.* Dynamic ElecTronic hEalth reCord deTection (DETECT) of individuals at risk of a first episode of psychosis: a case-control development and validation study. *Lancet Digit Heal* 2020;**2**:e229–39. doi:10.1016/S2589-7500(20)30024-8

27 Jackson RG, Patel R, Jayatilleke 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**. doi:10.1136/bmjopen-2016-012012

28 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. doi:10.1136/bmjopen-2020-042949

29 Holderness E, Cawkwell P, Bolton K, *et al.* 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. doi:10.18653/v1/w19-1915

30 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 Inf* 2020;**8**:e17376. doi:10.2196/17376

31 Ford T, Mansfield KL, Markham S, *et al.* The challenges and opportunities of mental health data sharing in the UK. *Lancet Digit Heal* 2021;**3**:e333–6. doi:10.1016/s2589-7500(21)00078-9

32 Holmusk. NeuroBlu: The Evidence Footprint. Building the future of behavioral health with Real-World Data and analytics. 2021.https://www.neuroblu.ai/

33 Mukherjee SS, Yu J, Won Y, *et al.* Natural Language Processing-Based Quantication of the Mental State of Psychiatric Patients. *Comput Psychiatry* 2020;**4**:76. doi:10.1162/cpsy_a_00030

34 Gersing K, Krishnan R. Clinical Computing: Clinical Management Research Information System (CRIS). *Psychiatr Serv* 2003;**54**:1199–200. doi:10.1176/appi.ps.54.9.1199

35 Response to the Goldacre Review - "Trusted Research Environments and Data Management". HDR UK. 2021.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* 2018;**2017**:1044–50.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977668/

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.https://doi.org/10.1007/978-3-030-70558-9_13

38 Patel R, Jayatilleke N, Broadbent M, *et al.* Negative symptoms in schizophrenia: a study in a large

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

clinical sample of patients using a novel automated method. *BMJ Open* 2015;**5**. doi:10.1136/bmjopen-2015-007619

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. doi:10.1093/schbul/sbx197

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**. doi:10.1136/bmjopen-2014-007504

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. doi:DOI: 10.1192/j.eurpsy.2020.39

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* Published Online First: 7 October 2020. doi:10.1093/schbul/sbaa126

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**:1178222618792860–1178222618792860. doi:10.1177/1178222618792860

44 Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. *Psychiatry (Edgmont)* 2007;**4**:28–37.https://pubmed.ncbi.nlm.nih.gov/20526405

45 *Diagnostic and statistical manual of mental disorders : DSM-IV*. Fourth edition. Washington, DC : American Psychiatric Association, [1994] ©1994 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 J Ment Sci* 1979;**134**:382–9. doi:10.1192/bjp.134.4.382

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**:38–52. doi:10.1017/S1092852914000601

49 Wu LT, 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. doi:10.1016/j.jpsychires.2013.08.022

50 Wu LT, 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. doi:10.1016/j.jpsychires.2011.06.012

51 Wu LT, 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. doi:10.1016/j.jpsychires.2012.12.009

52 Lee TS, Assam PN, Gersing KR, *et al.* The effectiveness of antidepressant monotherapy in a naturalistic outpatient setting. *Prim Care Companion J Clin Psychiatry* 2012;**14**. doi:10.4088/PCC.12m01364

53 Gersing KR, Sheehan JJ, Burchett B, *et al.* Use of augmentation agents for treating depression: Analysis of a psychiatric electronicmedical record data set. *Psychiatr Serv* 2014;**65**:1062–5. doi:10.1176/appi.ps.201300288

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. doi:10.1007/s11126-015-9389-8

55 Anisha Balani Sheryl Ker, Jennifer Hsu, Dustin Demoss, Miguel E. Rentería Sankha S. Mukherjee AJR. T4. Buprenorphine treatment retention in individuals with opioid use disorder: insights from real-world dataset. In: *American Society of Clinical Psychopharmacology*. 2021. https://pmg.joynadmin.org/documents/1005/60ad2955ef0d26cf5e684539.pdf#page=181

56 Mane KK, Owen P, Schmitt C, *et al.* Visual analytics to optimize patient-population evidence delivery

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

for personalized care. In: *2013 ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics, ACM-BCB 2013*. Association for Computing Machinery 2013. 484–8. doi:10.1145/2506583.2506608

57 Akushevich I, Kravchenko J, Gersing K, *et al.* 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. doi:10.1145/2506583.2506618

58 Pan JJ, 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. doi:10.1177/1740774508100983

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. doi:10.1016/j.jbi.2011.09.003

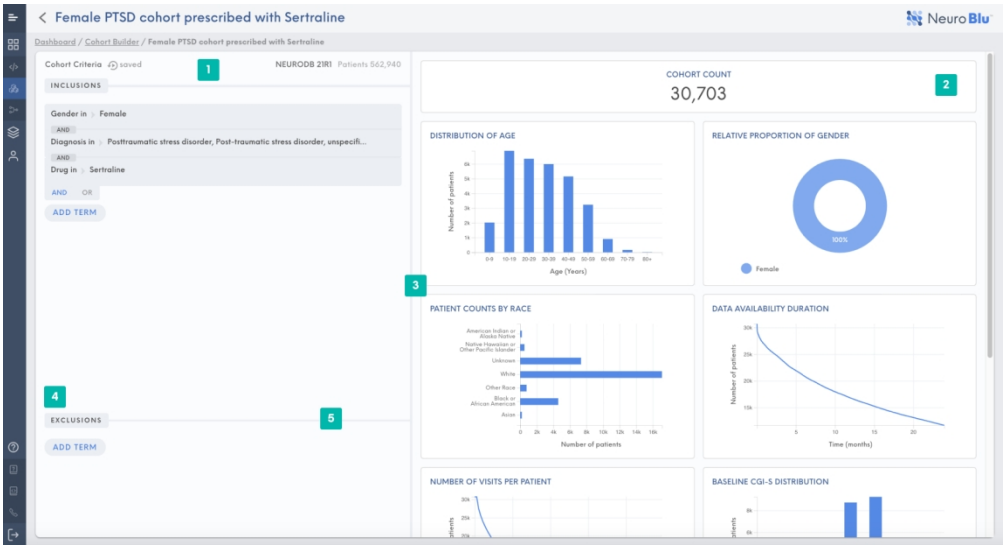
60 Sankha Mukherjee Yida Won, Mary McClay, Lu Wang, John Rush, Joydeep Sarkar JW. W35. Natural language processing-based quantification of the mental state of psychiatric patients. In: *American Society of Clinical Psychopharmacology*. 2021. https://pmg.joynadmin.org/documents/1005/60ad2955ef0d26cf5e684539.pdf#page=151

61 Hahn T, Nierenberg AA, Whitfield-Gabrieli S. Predictive analytics in mental health: applications, guidelines, challenges and perspectives. *Mol Psychiatry* 2017;**22**:37–43. doi:10.1038/mp.2016.201

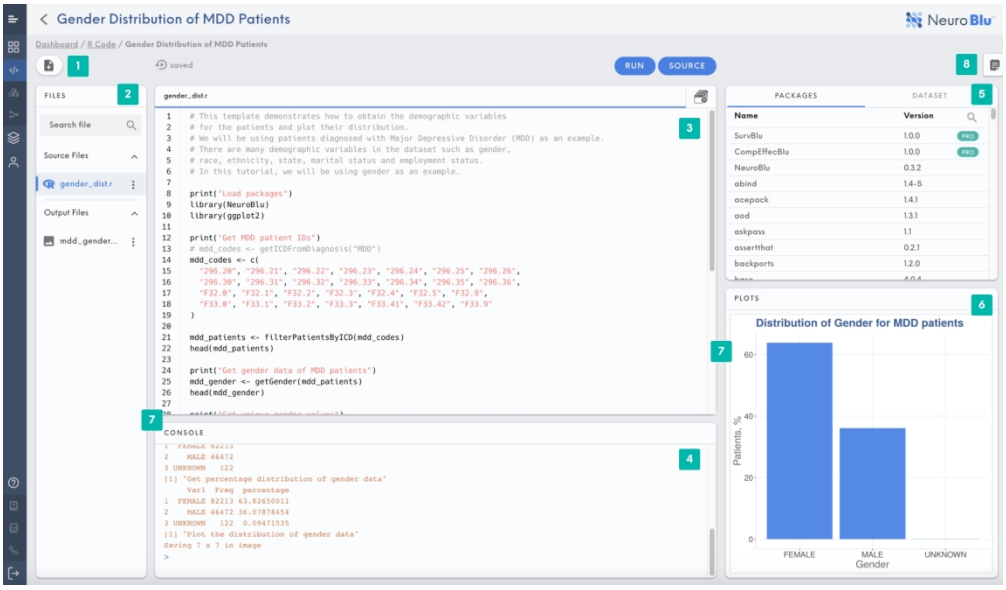
62 Holmusk. NeuroBlu Award 2021. 2021.https://www.holmusk.com/news/neuroblu-award

For peer review only

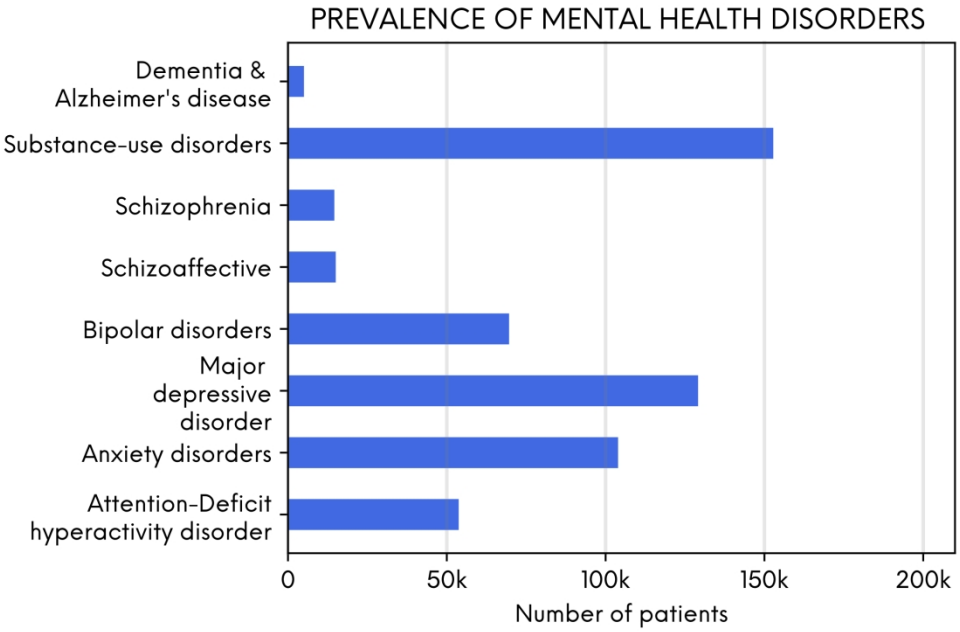
For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml



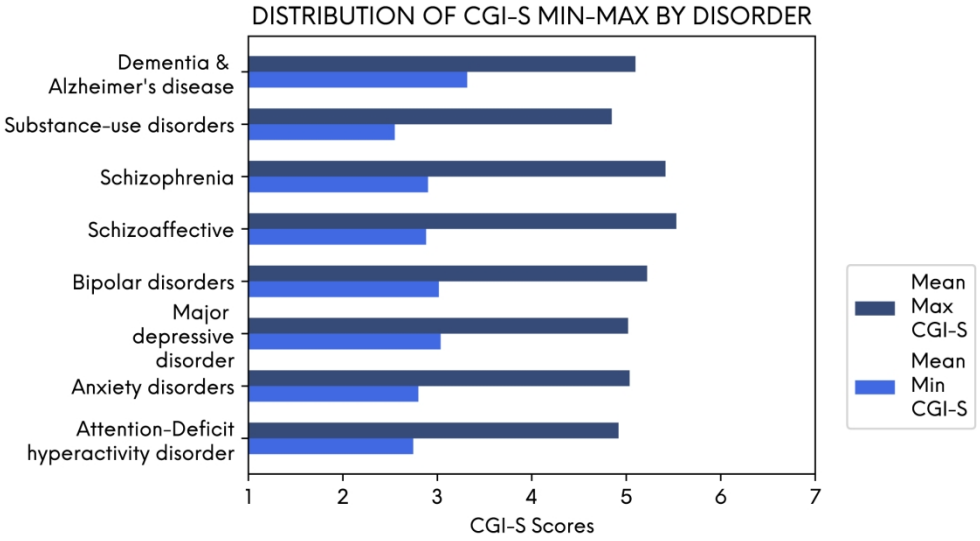
NeuroBlu Cohort Builder illustrating the population characteristics of female patients diagnosed with Post Traumatic Stress Disorder who have received sertraline



NeuroBlu R Code Engine which 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

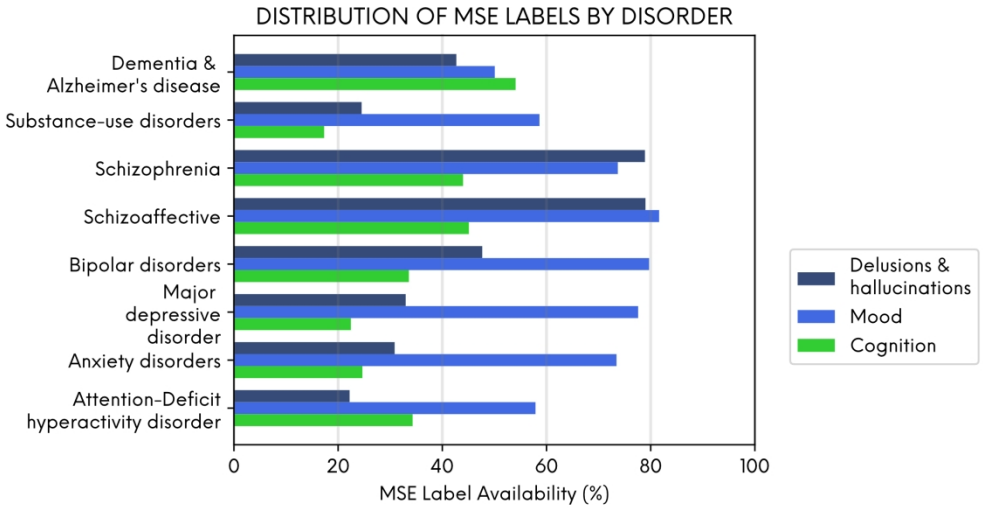


Prevalence of mental disorders in the NeuroBlu dataset



Mean maximum and mean minimum CGI-S score by diagnosis

177x99mm (300 x 300 DPI)



Percentage of patients with at least one NLP-derived MSE feature by diagnosis

Online supplementary material

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare real-world data analytics

Table of contents

Table of contents 1

1. NeuroBlu Common Data Model (CDM) pipeline 3

(i) Structural Normalization 3

(ii) Semantic Normalization 4

(iii) Data Validation 4

2. Data variables 5

(i) Clinical Data Tables 5

person 6

observation_period 8

visit_occurrence 9

condition_occurrence 10

drug_exposure 11

measurement 12

observation 14

(ii) Family History 17

(iii) Derived Measurements 18

(iv) Mental State Examination 23

eTable 1: Description of U.S. clinics contributing EHR data to NeuroBlu 31

eTable 2: NeuroBlu dataset race distribution 32

eTable 3: NeuroBlu dataset age distribution 32

eTable 4: NeuroBlu dataset diagnosis distribution 32

eTable 5: Gender distribution by diagnosis 33

eTable 6: ICD-9/ICD-10 code definitions for diagnostic category 34

eTable 7: U.S. State of Residence 65

eTable 8: Mean CGI-S score by diagnosis 67

eTable 9: Mean maximum and minimum CGI-S score by diagnosis 68

eTable 10: Number of patients with at least one hospital visit by diagnosis 69

eTable 11: SAV features included in MSE NLP analysis (eTable 12) 70

eTable 12: Number of patients with at least one MSE feature by diagnosis 71

eFigure 1: MindLinc EHR data de-identification and harmonization pipeline 72

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

eFigure 2: OMOP CDM schematic 73

eFigure 3: NeuroBlu Category Mapper illustrating the classification of individual drugs into different types of medication 74

eFigure 4: NeuroBlu Data Explorer which provides a summary of the dataset and allows the top 50 rows of each table to be viewed 75

eFigure 5: NeuroBlu dataset race distribution 76

eFigure 6: NeuroBlu dataset age distribution 77

eFigure 7: Gender distribution by diagnosis 78

eFigure 8: U.S. State of Residence 79

eFigure 9: Mean CGI-S score by diagnosis 80

eFigure 10: Percentage of patients with at least one hospital visit by diagnosis 81

eFigure 11: Distribution of number of visits per patient 82

eFigure 12: Distribution of duration of follow-up 83

For peer review only

Online supplementary material

1. NeuroBlu Common Data Model (CDM) pipeline

The NeuroBlu Data Model is based on the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) developed by the Observation Health Data Sciences and Informatics (OHDSI) community (eFigure 1). The model enables data contained within disparate observational databases to be transformed into a common format (the "data model") as well as a common representation (terminologies, vocabularies, coding schemes). This allows systematic analyses to be performed using a library of standard analytical routines that have been written based on the OMOP CDM.

The NeuroBlu Data Model includes the following modifications of the OMOP CDM to better support the representation of mental healthcare data:

1. Representing ICD9/ICD10 diagnostic codes in the CONDITION_OCCURRENCE table rather than the specified SNOMED concepts, as users typically have more familiarity with the ICD code sets than the SNOMED code set.
2. Adding the current value of mutable demographic attributes such as marital status, education, and occupation to the PERSON to allow users to access this information without needing to join a separate table.
3. Adding custom tables to support derived data and other data elements that do not have a natural home in the OMOP CDM. This includes derived measurements like smoothened CGI-S scores, imputed MADRS scores, and Mental State Examination (MSE) data obtained through natural language processing (NLP).

Once the MindLinc data are de-identified, transferred, and loaded into the relational PostgreSQL database, the remaining data integration processes required to ensure the data are ready for analysis within the NeuroBlu dataset include structural normalization, semantic normalization, and data validation.

(i) Structural Normalization

Structural normalization is the process of transforming the data into a single schema or layout. For example, to examine patient demographic information in a consistent way, regardless of how it was stored in the originating EHRs, the various patient and member demographic data are organized into a single, commonly structured PERSON table. This requires establishing and implementing various transformation rules that get applied to the data, such as those described below:

1. **Source-to-Target Mapping Logic:** this refers to the conveyance of information from a specific place in the source data to its related place in the target data model. For example, the MindLinc EHR source data contains a BACKGROUND table to record patient demographic attributes, such as RACE and SEX. In order for these attributes to end up in the appropriate place in the NeuroBlu Data Model, they need to be mapped from the relevant source field to the associated target field. In this example, the information in RACE and SEX MindLinc database would end up in the RACE_SOURCE_VALUE and GENDER_SOURCE_VALUE fields of the PERSON table in the NeuroBlu database.
2. **Inclusion/Exclusion Logic:** this refers to the criteria applied to the data in order for them to be included in the NeuroBlu Data Model. For example, information from the MindLinc source data that is marked as "entered in error" or "erroneous" brings into question whether it is truly evidence from real world data and in turn is excluded from the NeuroBlu Data Model.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

3. **De-duplication Logic:** this refers to the removal of duplicative or outdated information, such that only up-to-date data are pulled into the NeuroBlu Data Model and are only pulled in once. Depending on the source data elements of interest and the mechanism by which updates are applied to them, the method for de-duplication will vary. In the case of the MindLinc data, a destructive refresh approach is taken across all data tables such that the most recently acquired data will always replace the older data.

(ii) Semantic Normalization

Semantic normalization is the process of translating the data to a single set of units and vocabulary. For example, in order to specifically examine female patients in the aforementioned PERSON table, the various ways the patient and member demographic data codify the concept of "female" need to be examined such that a translation between the source codes (e.g. "F", "FEM", "FEMALE", "0", "396123", etc.) and the NeuroBlu standard concept ("FEMALE") can be created and applied to the data. These translations are generated through a combination of direct-mapping, text-string-matching, and NLP based on a thorough examination of the native codes, their descriptions, any supporting information provided about them (e.g. reference ranges for scores, generic name for medications, etc.), and, if necessary, expert input from the healthcare provider organization they are originating from.

A number of industry standard reference code sets are leveraged as part of the semantic normalization process for the MindLinc data, including RxNorm for drug mapping and ICD9/ICD10 for diagnostic code mapping. The target concept sets for all normalized data elements are the Standardized Vocabulary provided as part of the OMOP CDM standard.

(iii) Data Validation

A core component of data acquisition and integration is ensuring that the resultant data set is of the highest quality possible, where quality is defined as a function of conformance, completeness, correctness, and consistency. In order to measure the quality of newly integrated data against these key dimensions, the NeuroBlu CDM pipeline employs an approach to data validation that uses the wealth of data available during the process to produce a composite measure of what certain key data points typically look like in different situations, and then evaluates the data against these expectations. Doing this efficiently requires validation of the data as they flow through the processing pipeline, which in turn necessitates a robust strategy for determining key quality indication metrics, generating those metrics in a systematic manner, and distilling those metrics into meaningful insights.

The Data Validation process works in concert with the above normalization processes to identify issues or inconsistencies in the data that need to be addressed and provide guidelines on how those various issues or inconsistencies should be handled. Erroneous data that can't be fixed at the source are either remedied or removed by NeuroBlu's Extract, Transform, Load (ETL) processes, while quirks or unexpected features of the data that are not necessarily errors are systematically logged.

The end result of these data integration processes is the creation of a single normalized and validated Common Data Model for use in the NeuroBlu trusted research environment.

For peer review only

Online supplementary material

2. Data variables

(i) Clinical Data Tables

The clinical data tables contain the core information about the clinical events that occurred longitudinally during valid Observation Periods for each person, as well as demographic information for the person. In addition to the standardized clinical data tables defined in the OMOP CDM, the NeuroBlu dataset also contains other tables that are not defined in the OMOP CDM including data on Family History, Derived Measurements and Mental State Examination and which are described in greater detail in sections (ii), (iii) and (iv). The tables form part of a relational database in which "_id" columns may be joined between tables to assemble clinical data at individual person level.

Table name	Type	Content
person	CDM	Socio-demographic information of each patient including year of birth, gender, race, location of residence, occupational status and education status.
observation_period	CDM	Time interval for which a patient has data recorded. Data from this table is used to filter cohorts according to minimum duration within the dataset to support longitudinal analyses.
visit_occurrence	CDM	Records of each visit and its type. Data may be obtained from this table on the number of contacts with the treating mental healthcare service and the type of contact (e.g. outpatient or inpatient).
condition_occurrence	CDM	Records of clinical diagnosis in the form of ICD-9-CM and ICD-10-CM codes. Data from this table is used to generate population cohorts based on primary mental disorder diagnosis and to obtain data on comorbid diagnoses.
drug_exposure	CDM	Records of any exposure to prescribed medications including the names of prescribed medications, start and stop dates, dose and frequency of administration.
measurement	CDM	Records of direct clinical measurements using clinician-rated structured rating scales.
observation	CDM	NLP-derived records of social factors which could contribute towards risk of developing a mental disorder and/or poorer clinical outcomes. This table can also be used to obtain social factor data as outcomes (e.g. social and occupational functioning)
family_history	Non-CDM	Provides recorded data on family history of mental disorders
derived_measurement	Non-CDM	Clinical measurements which have been transformed or derived from recorded CGI-S scores.
derived_mse	Non-CDM	NLP-derived mental state examination (MSE) data.
CDM: OMOP Common Data Model Non-CDM: NeuroBlu customized data structure (not defined in OMOP)		

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

person

The person table contains records that uniquely identify each person in the dataset. This table also contains demographic information about each individual (excluding data removed by the de-identification process). All other clinical data tables refer to this table using the person_id field which is the unique record identifier for each individual represented in the dataset.

Columns

Field	Type	Description
person_id	int8	A unique identifier for each person. This value is used to join onto other tables in NeuroBlu to obtain clinical data at individual person level.
gender_concept_id	int4	A foreign key that refers to an identifier in the concept table for the unique gender of the person, belonging to the 'Gender' Vocabulary. Gender data are obtained by joining on this value (see below).
year_of_birth	int4	The year of birth of the person.
race_concept_id	int4	A foreign key that refers to an identifier in the concept table for the unique race of the person, belonging to the 'Race' Vocabulary. Race data are obtained by joining on this value (see below).
ethnicity_concept_id	int4	A foreign key that refers to an identifier in the concept table for the unique ethnicity of the person, belonging to the 'Ethnicity' Vocabulary. Ethnicity data are obtained by joining on this value (see below).
zip_code_source_value	varchar	3-digit ZIP code of the person's location.
city	varchar	City name based on 3-digit ZIP code.
state	varchar	State name based on 3-digit ZIP code.
marital_status	varchar	Marital status of the person.
occupation	varchar	Employment status of the person.
education	int2	Number of years of formal education. When education_source_value is not available, defaults to 0
education_source_value	varchar	Source value for education. Not always available.

Conventions

No.	Convention Description
1	All tables representing patient-related domains have a foreign-key reference to the person_id field in the person table.
2	Each person record has associated demographic attributes which are assumed to be constant for the patient throughout the course of their periods of observation. For example, the location or gender is expected to have a unique value per person, even though in life these data may change over time.
3	The month and day of birth for a person is not recorded. Therefore, value defaults to Jan 1st.
4	Ethnicity and Race in the OMOP CDM follow the 1997 Office of Management and Budget (OMB) standards for data on race and ethnicity.

For peer review only

Online supplementary material

Gender Vocabulary

concept_id	concept_name
8507	MALE
8532	FEMALE
8551	UNKNOWN

Race Vocabulary

concept_id	concept_name
8515	Asian
8516	Black or African American
8522	Other Race
8527	White
8552	Unknown
8557	Native Hawaiian or Other Pacific Islander
8657	American Indian or Alaska Native

Ethnicity Vocabulary

concept_id	concept_name
759814	Ethnic group unknown
38003563	Hispanic or Latino
38003564	Not Hispanic or Latino

Marital Status Vocabulary

marital_status
death of spouse
divorced
engaged
married
separated
single
together
unknown

Occupation Vocabulary

occupation
Disabled
Employed
Retired
Student
Unemployed
Unknown

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

observation_period

The observation_period table contains records which uniquely define the spans of time for which a patient is expected to have clinical events recorded within the source systems, even if no events in fact are recorded (healthy patient with no healthcare interactions).

Columns

Field	Type	Description
observation_period_id	serial	A unique identifier for each observation period.
person_id	int8	A foreign key identifier to the person for whom the observation period is defined. The demographic details of that person are stored in the person table.
observation_period_start_date	date	The start date of the observation period for which data are available from the data source.
observation_period_end_date	date	The end date of the observation period for which data are available from the data source.
period_type_concept_id	int4	A foreign key identifier to the predefined Concept in the Standardized Vocabularies reflecting the source of the observation period information, belonging to the 'Obs Period Type' Vocabulary.

Conventions

No.	Convention Description
1	Each person has to have at least one observation period.
2	One person may have one or more disjoint observation periods, during which times analyses may assume that clinical events would be captured if observed.
3	Each person can have more than one valid observation period record, but no two observation periods can overlap in time for a given person.
4	As a general assumption, during an observation period any clinical event that happens to the patient is expected to be recorded. Conversely, the absence of data indicates that no clinical events occurred to the patient.
5	Both the start date and the end date of the clinical event has to be between observation_period_start_date and observation_period_end_date.
6	For EHR data, the observation period cannot be determined explicitly, because patients usually do not announce their departure from a certain healthcare provider. The extract, transform, load (ETL) process will have to apply some heuristic to make a reasonable estimate on what the observation_period should be.

For peer review only

Online supplementary material

visit_occurrence

The visit_occurrence table contains the spans of time a patient continuously receives medical services from one or more providers at a care site in a given setting within the health care system.

Columns

Field	Type	Description
visit_occurrence_id	int8	A unique identifier for each person's visit or encounter at a healthcare provider.
person_id	int8	A foreign key identifier to the person for whom the visit is recorded. The demographic details of that person are stored in the person table.
visit_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies belonging to the 'Visit' Domain.
visit_start_datetime	timestamp	The date and time when the visit started. The time is not captured in the source and defaults to 00:00:00.
visit_end_datetime	timestamp	The date and time when the visit ended. Not recorded in the source. Takes the same value as visit_start_datetime.
visit_type_concept_id	int4	A foreign key identifier to the predefined Concept in the Standardized Vocabularies reflecting the source data from which the visit record is derived from, belonging to the 'Visit Type' Vocabulary. Visit type data are obtained by joining on this value (see below).
provider_id	int8	A foreign key to the primary care provider the person is seeing.
care_site_id	int8	A foreign key to the site of primary care.
visit_source_value	varchar	Source value for visit_concept_id.

Conventions

No.	Convention Description
1	A visit occurrence is recorded for each engagement of a patient with the healthcare system. It reflects any interaction with the healthcare service, not just in person clinical interactions.
2	Valid Visit Concepts belong to the 'Visit' domain.
3	At any one day, there could be more than one visit.
4	Visit times and visit end dates are not captured in the source. They resort to 00:00:00 and visit start dates, respectively.

Visit types

The table below describes the different visit types documented in the visit_occurrence table.

concept_id	concept_name
262	Emergency Room and Inpatient Visit
9201	Inpatient Visit
9202	Outpatient Visit
32693	Health examination
581458	Pharmacy visit
581477	Office Visit
42898160	Non-hospital institution Visit

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

condition_occurrence

The condition_occurrence table contains the presence of a disease or medical condition recorded as a diagnosis according to the International Classification of Disease (ICD), which is either observed by a provider or reported by the patient. In the NeuroBlu dataset, diagnoses are recorded as ICD-9-CM and ICD-10-CM codes in the source.

Columns

Field	Type	Description
condition_occurrence_id	serial	A unique identifier for each condition occurrence event.
person_id	int8	A foreign key identifier to the person for whom the condition is recorded. The demographic details of that person are stored in the person table.
condition_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies belonging to the 'Condition' Domain. Joining on this value extracts the corresponding ICD-9-CM or ICD-10-CM diagnostic code.
condition_start_datetime	timestamp	The date and time when the instance of the condition is recorded. The time is not captured in the source and defaults to 00:00:00.
provider_id	int8	A foreign key to the primary care provider the person is seeing.
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the condition was diagnosed.
condition_source_value	varchar(255)	Source value for condition_concept_id.

Conventions

No.	Convention Description
1	Past diagnoses ('history of') are not recorded in the condition_occurrence table. Instead, they are listed in the observation table.
2	The exact start date and end date for a condition is not recorded explicitly. Rather, the start date contains the date when the diagnosis was made, and the end date is left empty.

For peer review only

Online supplementary material

drug_exposure

The drug_exposure table contains records about the utilization of a prescribed medication when ingested or introduced to the body. Prescribed medications are inferred from clinical events associated with orders, prescriptions written, pharmacy dispensing, procedural administrations, and other patient-reported information.

Columns

Field	Type	Description
drug_exposure_id	serial	A unique identifier for each condition exposure event.
person_id	int8	A foreign key identifier to the person for who is subjected to the drug. The demographic details of that person are stored in the person table.
drug_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies belonging to the 'Drug' Domain. Joining on this value extracts the name of the prescribed medication.
drug_exposure_start_datetime	timestamp	The start date and time for the current instance of drug utilization. Valid entries include a start date of a prescription or the date on which a drug was administered. The time is not captured in the source and defaults to 00:00:00.
drug_exposure_end_datetime	timestamp	The end date and time for the current instance of drug utilization. It could be a known or an inferred date and denotes the last day at which the patient was still exposed to the drug during an instance of drug utilization.. The time is not captured in the source and defaults to 00:00:00.
quantity	numeric	Total quantity of drug exposed, in mg. Not always available.
days_supply	int4	The number of days of drug exposure. Not always available.
provider_id	int8	A foreign key to the primary care provider the person is seeing.
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the drug exposure was initiated.
drug_source_value	varchar(50)	Source value for drug_concept_id.

Conventions

No.	Convention Description
1	Valid Concepts for the drug_concept_id field belong to the 'Drug' domain. They are based on Ingredient Concept Class within the RxNorm Vocabulary.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

measurement

The measurement table contains records of structured values (numerical or categorical) obtained through systematic and standardized examination or testing of a person. In the NeuroBlu dataset, the table contains measurements of:

1. Clinical Global Impression - Severity (CGI-S)
2. Clinical Global Impression - Improvement (CGI-I)
3. Global Assessment of Functioning (GAF)
4. Montgomery-Asberg Depression Rating Scale (MADRS)

Columns

Field	Type	Description
measurement_id	serial	A unique identifier for each measurement.
person_id	int8	A foreign key identifier to the person for whom the measurement is recorded. The demographic details of that person are stored in the person table.
measurement_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies. Joining on this value specified which clinical measurement is being recorded.
measurement_datetime	timestamp	The date and time of the measurement. The time is not captured in the source and defaults to 00: 00: 00.
measurement_type_concept_id	int4	A foreign key to the predefined Concept in the Standardized Vocabularies reflecting the provenance from where the Measurement record was recorded. These belong to the 'Meas Type' vocabulary.
operator_concept_id	int4	A foreign key identifier to the predefined Concept in the Standardized Vocabularies reflecting the mathematical operator that is applied to the value_as_number. Operators are <, < = , = , > = , > and these concepts belong to the 'Meas Value Operator' domain.
value_as_number	numeric	A measurement result where the result is expressed as a numeric value.
provider_id	int8	A foreign key to the primary care provider the person is seeing.
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the measurement was recorded.
measurement_source_value	varchar	Source value for measurement_concept_id.
measurement_source_concept_id	int4	A foreign key to a Concept in the Standardized Vocabularies that refers to the code used in the source.
unit_source_value	varchar	Source value for the unit.
value_source_value	varchar	Source value associated with value_as_number column.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

custom2_str	varchar	Custom field to indicate which measurement the record is for.

Conventions

No.	Convention Description
1	Measurements are stored as attribute value pairs, with the attribute as the measurement_concept_id and the value representing the result (value_as_number).
2	The operator_concept_id is optionally given for relative Measurements where the precise value is not available but its relation to a certain benchmarking value is available. For example, this can be used for minimal detection thresholds of a test.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

observation

The observation table contains records of clinical facts about a related to medical history, the stated need for certain treatment, social circumstances, lifestyle choices, healthcare utilization patterns, etc. These records are obtained during examination or procedure and differ from Measurements in that they do not require a standardized test to generate clinical fact.

In NeuroBlu, the observation table contains the social history in the form of external environmental stressors.

Columns

Field	Type	Description
observation_period_id	serial	A unique identifier for each observation.
person_id	int8	A foreign key identifier to the person for whom the observation period is defined. The demographic details of that person are stored in the person table.
observation_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies.
observation_datetime	timestamp	The date and time of the observation. The time is not captured in the source and defaults to 00: 00: 00.
observation_type_concept_id	int4	A foreign key to the predefined Concept in the Standardized Vocabularies reflecting the provenance from where the Observation record was recorded. These belong to the 'Observation Type' vocabulary.
value_as_string	varchar(255)	Categorical value of the Result of the Observation, if applicable and available.
value_as_concept_id	int4	A foreign key to an observation result stored as a Concept ID. This is applicable to observations where the result can be expressed as a Standard Concept from the Standardized Vocabularies (e.g., positive/negative, present/absent, low/high, etc.).
qualifier_concept_id	int4	Contains all attributes specifying the clinical fact further, such as degrees, severities, drug-drug interaction alerts
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the measurement was recorded.
observation_source_value	varchar(255)	The observation code (Eg: ICD10) as it appears in the source data.
qualifier_source_value	varchar(50)	The source value associated with a qualifier to characterize the observation.

For peer review only

Online supplementary material

Conventions

No.	Convention Description
1	Observations can be stored as attribute value pairs, with the attribute as the Observation Concept and the value representing the clinical fact.
2	Valid Observation Concepts are not enforced to be from any domain though they still should be Standard Concepts.

External Stressors data derived using NLP

In the NeuroBlu dataset, the raw data for external environmental stressors (or social history) is a mixture of structured values and custom, user-defined free text. These values need to be normalized to structured values before they are used for downstream analytics. To achieve this, we have applied Natural Language Processing (NLP) to map the unstructured free text to structured labels.

The NLP model maps the semi-structured stessor data to one of the possible 61 labels (which can be categorized into 8 categories). Each label is also associated with a polarity (positive, negative or neutral) as described in the table below.

Category	Stressor Label	Polarity
Academic	Discord with teachers and classmates	Negative
Academic	Good achievement in school	Positive
Academic	Illiteracy and low-level literacy	Negative
Academic	Suspension/expulsion	Negative
Academic	Underachievement in school	Negative
Environment	Abusive environment at home	Negative
Environment	Chaotic atmosphere / stressful home environment	Negative
Environment	Cultural conflicts/issue	Negative
Environment	Exposure to Disaster, War, or Other Hostilities	Negative
Environment	Inadequate care (history from caregiver or patient)	Negative
Environment	Migration/ Acculturation Difficulty	Negative
Environment	Problem Related to Living Alone	Negative
Environment	Stable living situation	Positive
Family	Anxiety about sick person in family, dependent relative needing care at home	Negative
Family	Bereavement	Negative
Family	Disruption of Family by Separation or Divorce	Negative
Family	Inadequate or distorted communication within family	Negative
Family	Isolated family, lack of family support	Negative
Family	Legal problem in the family	Negative
Family	Parenting situation	Negative
Family	Problems in parent-child relationships and in-laws	Negative
Family	Sibling Relational Problem	Negative
Family	Substance abuse in the family	Negative
Family	Supportive family, love and respect in the family	Positive
Financial	Extreme Poverty/ Low income	Negative
Financial	Financial stability	Positive
Financial	Inadequate health/social Insurance or Welfare Support	Negative
Financial	Poor money management / no savings	Negative
Financial	Problems related to loan / creditors	Negative
Housing	Discord With Neighbor	Negative

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Housing	Homelessness	Negative
Housing	Inadequate housing	Negative
Housing	Isolated dwelling (Lack of transportation)	Negative
Housing	Unstable housing arrangement	Negative
Legal	Arrest, Imprisonment and other incarceration	Negative
Legal	Child custody or support proceedings	Negative
Legal	Litigation / legal proceedings	Negative
Legal	Offense / charges	Negative
Legal	Problem with legal personnel	Negative
Legal	Restrain order	Negative
Occupational	Change of job	Neutral
Occupational	Fired	Negative
Occupational	Job dissatisfaction	Negative
Occupational	Job satisfaction	Positive
Occupational	Occupational stress	Negative
Occupational	Threat of job loss, Underperform at work	Negative
Occupational	Unemployment	Negative
Occupational	Unstable job (changing jobs often)	Negative
Personal	Emotional abuse	Negative
Personal	Involuntary treatment	Negative
Personal	Major life events	Negative
Personal	Physical abuse (spouse)	Negative
Personal	Relationship Distress With Spouse or Intimate Partner	Negative
Personal	Sense of meaning and support from belief	Positive
Personal	Sexual abuse (spouse)	Negative
Personal	Target of Adverse Discrimination	Negative
Personal	Victim of crime	Negative
Social	Conflictual social relationships	Negative
Social	Inadequate social support	Negative
Social	Lack of social circles	Negative
Social	Supportive friends	Positive

Precision statistics for the NLP classifiers provided as the median metrics of each stressor category are provided in the table below.

Category	AUROC	Recall	Precision	F1
Academic	0.999	0.667	1.000	0.800
Environment	0.999	0.923	0.800	0.857
Family	0.984	0.519	0.759	0.619
Financial	0.998	0.500	0.833	0.667
Housing	0.993	0.417	0.633	0.473
Legal	0.985	0.708	0.813	0.745
Occupational	0.993	0.500	0.895	0.664
Social	0.990	0.500	0.833	0.625

For peer review only

Online supplementary material

(ii) Family History

The family_history table contains data on recorded diagnoses of mental disorders affecting patients' relatives in the form of ICD-10-CM codes, along with the degree of relationship relative to the patient.

Columns

Field	Type	Description
family_history_id	serial	A unique identifier for each family history record.
person_id	int8	A foreign key identifier to the person for whom the MSE is recorded. The demographic details of that person are stored in the person table.
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the data was recorded.
condition_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies belonging to the 'Condition' Domain. Joining on this value extracts the corresponding ICD-9-CM or ICD-10-CM diagnostic code.
condition_source_value	varchar(255)	Source value for condition_concept_id.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

(iii) Derived Measurements

The derived_measurement table contains records of structured values (numerical or categorical) obtained through deriving recorded measurements in the measurement table. In the NeuroBlu dataset, the table contains measurements of:

1. Smoothened CGI-S
2. Derived MADRS

Columns

Field	Type	Description
derived_measurement_id	serial	A unique identifier for each derived measurement.
person_id	int8	A foreign key identifier to the person for whom the measurement is recorded. The demographic details of that person are stored in the person table.
derived_measurement_value	float4	The value of the derived measurement.
derived_measurement_date	date	The date of the derived measurement.
range_low	numeric	The lower limit of the normal range of the measurement result. The lower range is assumed to be of the same unit of measure as the measurement value.
range_high	numeric	The upper limit of the normal range of the measurement result. The upper range is assumed to be of the same unit of measure as the measurement value.
provider_id	int8	A foreign key to the primary care provider the person is seeing.
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the measurement was recorded.
unit_source_value	varchar(50)	Source value for the unit.
custom2_str	varchar(255)	Custom field to indicate which measurement the record is for (i.e. Derived MADRS or Smoothened CGI-S)

Conventions

No.	Convention Description
1	Measurements are stored as attribute value pairs, with the attribute as the custom2_str and the value representing the result (derived_measurement_value).

For peer review only

Online supplementary material

Smoothened CGI-S

In the NeuroBlu dataset, Clinical Global Impression-Severity (CGI-S) measurements are found to be noisy, both the severity and the improvement scales. The values recorded fluctuate across measurements. Sometimes a visit doesn't have any CGI-S measurement, and sometimes multiple CGI-S measurements are recorded in a single day.

To tackle this problem, we provide smoothened values of CGI-S. Raw measurement values are smoothened and interpolated on days with visit but no CGI-S measurements (the visit is recorded in visit_occurrence, but no CGI-S is measured on that day). These smoothened and interpolated values are available in the derived_measurement table.

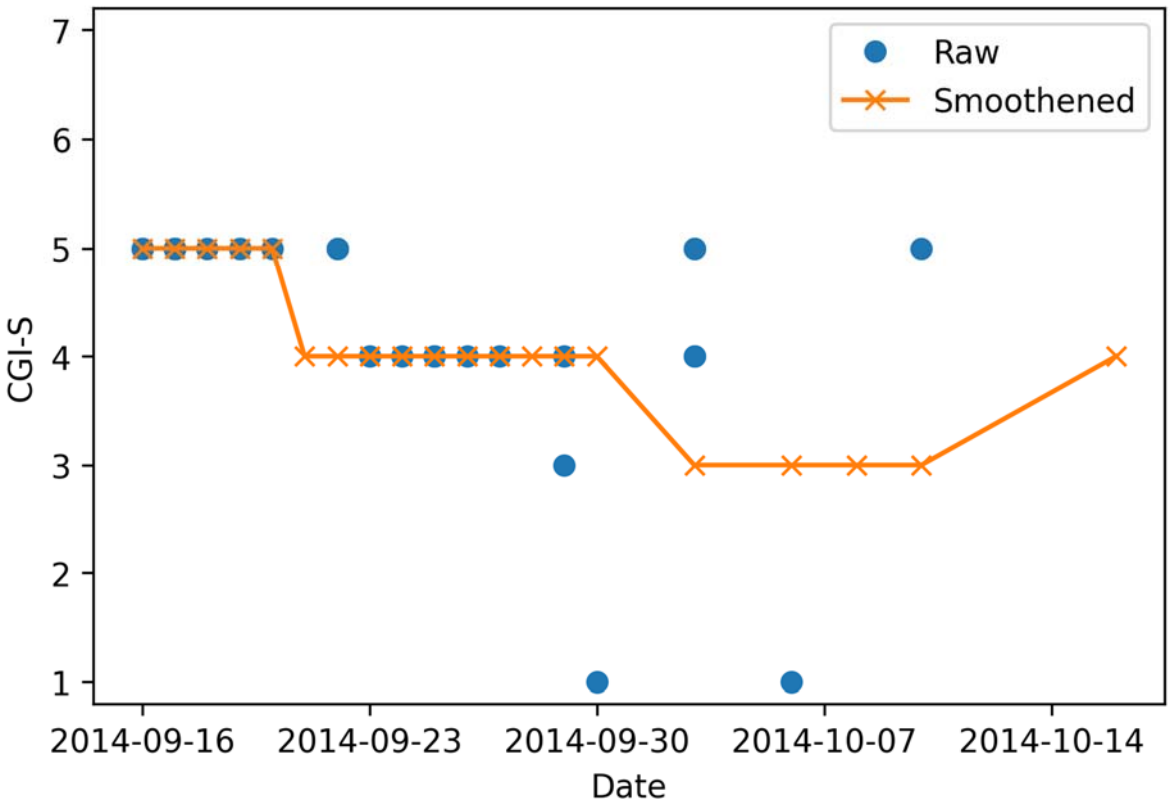
The smoothing and interpolation is done on a per-patient basis using exponential-weighted smoothing, based on the following equation:

$$s_i = \sum_{j \in J} w_{i,j} c_j$$

$$w_{i,j} = \frac{\exp(-d_{i,j}^2/(2\sigma^2))}{\sum_{j \in J} \exp(-d_{i,j}^2/(2\sigma^2))}$$

where *s_i* is the smoothened CGI-S at visit *i*, *J* is the set of all visits with CGI-S measurements, *c_j* is the measured raw CGI-S values at visit *j*, *d_i,j* is the time difference between visit *i* and *j*, in terms of days, and *σ* is the smoothing parameter, equals to 5 in our case. The equation works for days with multiple CGI measurements and days without CGI-S measurements.

The following figure shows how the smoothened CGI-S looks for an example patient.



Online supplementary material

MADRS Scores derived using CGI-S data

The Montgomery–Asberg Depression Rating Scale (MADRS) is a clinician-recorded rating scale for symptoms of depression. It comprises of 10 individual scores which are summed to yield a final score which ranges from 0 to 60. However, unlike the Clinical Global Impression-Severity (CGI-S), the NeuroBlu dataset has few documented records of MADRS scores. However, it is also one of the most widely used scores in psychiatry. We have observed that there is a fairly high correlation between the CGI-S and the final MADRS score. For this reason mapping functions have been derived that may be used to map the CGI-S score to the MADRS score.

In the event an analysis requires MADRS data, we provide values of MADRS derived from CGI-S at every timepoint with CGI-S measurement. These smoothened and interpolated values are available in the derived_measurement table.

All encounters in which both CGI-S and MADRS scores have been simultaneously measured are extracted from the dataset. In total, there are approximately 3101 such encounters. Both measurements are strongly correlated with a correlation coefficient of 0.80. This supports the possibility of mapping CGI-S to MADRS.

Upon manual inspection of the distribution of the MADRS scores for each value of CGI-S, we observed that MADRS scores have asymmetric distribution shapes for CGI-S values in the end of the spectrum, and wide distribution shapes for CGI-S values around the middle of the spectrum. For this reason, we feel that a simple linear model may not provide good confidence interval for the derived MADRS from CGI-S. We conclude that MADRS score follows the four-parameter beta distribution Beta(a, b, 0, 60), where the parameters a, b is a function of the CGI-S score, and the value is bounded between 0 and 60. Then, the values of a, b are estimated for every possible value of CGI-S using the method of moments. The derived MADRS for each CGI-S score takes value of the distribution mean, i.e. $60*a/(a+b)$. To gauge the goodness-of-fit of this approach, the R^2 value is computed and is found to be 0.66.

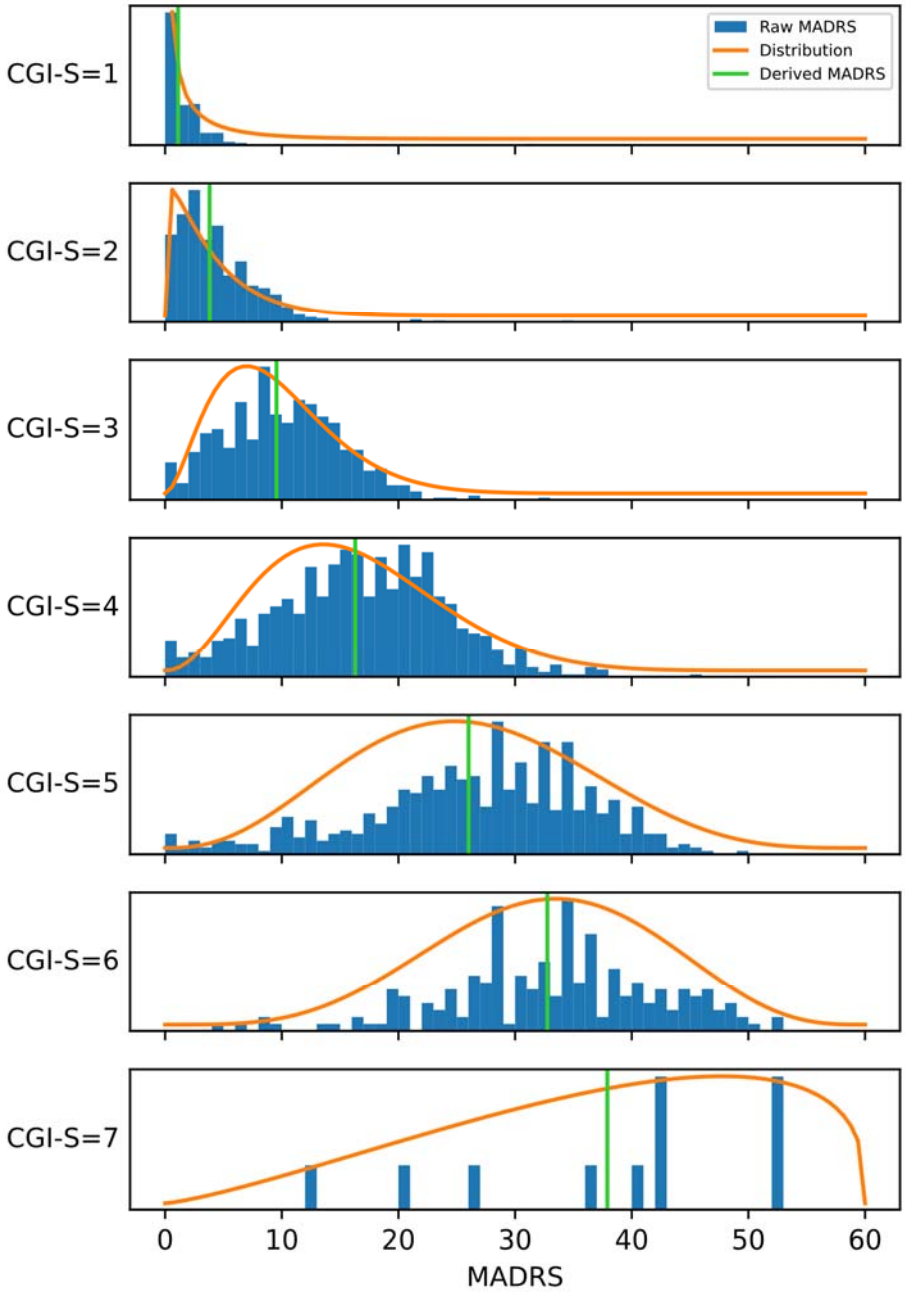
The estimated values of a and b along with the derived MADRS value for each CGI-S score are summarized in the following table.

CGI-S	a	b	Derived MADRS	95% confidence interval
1	0.1874	9.9098	1.1	0.9 to 1.3
2	1.1070	16.3081	3.8	3.6 to 4.0
3	2.8634	15.1325	9.5	9.3 to 9.8
4	3.2716	8.7725	16.3	16.0 to 16.6
5	3.7353	4.8806	26.0	25.5 to 26.6
6	5.1854	4.3100	32.8	31.9 to 33.6
7	2.2803	1.3288	37.9	32.7 to 43.1

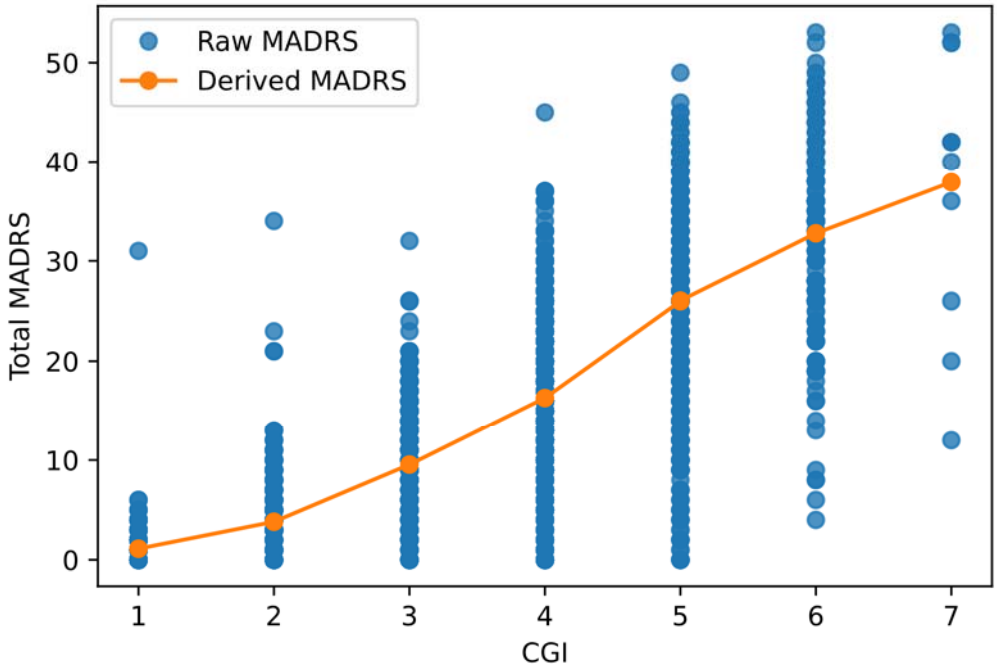
The distributions of MADRS score for each value of CGI-S are shown in the following figures. The first figure shows the histogram of MADRS scores for each value of CGI-S and the shape of the fitted beta distributions. The second figure shows the relationship between derived MADRS and CGI-S on top of a scatter plot of raw MADRS versus raw CGI-S.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material



Online supplementary material



er review only

(iv) Mental State Examination

The derived_mse table contains the cleaned form of the mental state examination (MSE) records. In the NeuroBlu dataset, the MSE data are represented as Status Assessment Vectors (SAV) which are further described subsequently.

Columns

Field	Type	Description
person_id	int8	A foreign key identifier to the person for whom the MSE is recorded. The demographic details of that person are stored in the person table.
measurement_date	date	The date of the derived MSE.
Days	int4	Number of days of measurement since the patient was first recorded in the dataset.
sav_i_cat_j	int4	i-th value of the SAV. Also correspond to the j-th value for cat category. Contains binary value. i takes value from 0 to 240.

Status Assessment Vectors

In the NeuroBlu dataset, the raw data for MSE is recorded in the form of semi-structured free text, making it difficult to analyze in their raw form.

To tackle this problem, we provide natural language processing (NLP) derived structured values of MSE. Raw MSE data for a single patient on a particular day are cleaned and passed through an NLP model that returns a 241-element vector form of MSE. This vector is called the Status Assessment Vector (SAV). SAV is a binary vector, where the value is 1 if the category-factor combination of the SAV is applicable for the patient. As an example, the 15th position in the SAV corresponds to whether a patient is responding to internal stimuli or not and value of 1 for this position means that the patient is responding to internal stimuli. Further details of the NLP pipeline are described in the following article:

Mukherjee, S.S., Yu, J., Won, Y., McClay, M.J., Wang, L., Rush, A.J. and Sarkar, J., 2020. Natural Language Processing-Based Quantification of the Mental State of Psychiatric Patients. *Computational Psychiatry*, 4, pp.76–106. DOI: http://doi.org/10.1162/cpsy_a_00030

The table below maps each element of the SAV to its location in the derived_mse table.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

SAV #	Category	Factor
0	Abnormal or psychotic thoughts	no issues
1	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (grandeur)
2	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (persecution)
3	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (religious)
4	Abnormal or psychotic thoughts	experiencing hallucinations (auditory)
5	Abnormal or psychotic thoughts	experiencing hallucinations (visual)
6	Abnormal or psychotic thoughts	experiencing hallucinations (tactile)
7	Abnormal or psychotic thoughts	History of delusions/abnormal thoughts
8	Abnormal or psychotic thoughts	History of hallucinations
9	Abnormal or psychotic thoughts	neutral/unable to categorize
10	Abnormal or psychotic thoughts	experiencing delusions (not specified)
11	Abnormal or psychotic thoughts	experiencing hallucinations (not specified)
12	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (obsessions)
13	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (paranoia)
14	Abnormal or psychotic thoughts	experiencing hallucinations (olfactory)
15	Abnormal or psychotic thoughts	responding to internal stimuli
16	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (sexual)
17	Affect	no issues/appropriate affect
18	Affect	neutral/unable to categorize
19	Affect	aggressive
20	Affect	reactive
21	Affect	irritable/angry
22	Affect	improved from last visit
23	Affect	issues with cognition
24	Affect	loss of energy/tiredness

For peer review only

Online supplementary material

25	Affect	blunted/restricted
26	Affect	anxious/fearful
27	Affect	tearful/depressed/sad
28	Affect	bright/positive
29	Affect	labile
30	Affect	intense
31	Affect	expansive/overly animated
32	Affect	generally odd or out of the ordinary/inappropriate
33	Affect	flooded
34	Appearance	Issues with Affect
35	Appearance	Sleep/Bed
36	Appearance	Issues with Eye Contact
37	Appearance	Not age appropriate
38	Appearance	Issues with general Health
39	Appearance	Issues with Behavior
40	Appearance	Positive Affect
41	Appearance	Positive Grooming/Hygiene
42	Appearance	Issues with Grooming/Hygiene
43	Appearance	Positive Health
44	Appearance	Issues with Weight
45	Appearance	Normal/no issues
46	Appearance	Hospital attire
47	Appearance	Positive dress
48	Appearance	Issues with dress
49	Appearance	Neutral
50	Appearance	Positive eye contact
51	Appearance	Appropriate Age
52	Appearance	Positive Behavior
53	Association	intact/no issues
54	Association	neutral/unable to categorize
55	Association	circumstantial
56	Association	loose
57	Association	tangential
58	Association	age appropriate
59	Association	improved
60	Association	declined
61	Association	disorganized
62	Association	general issues
63	Attention/ Concentration	intact/no issues
64	Attention/ Concentration	neutral/unable to categorize

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

65	Attention/ Concentration	general issues
66	Attention/ Concentration	improved
67	Attention/ Concentration	declined
68	Attention/ Concentration	age appropriate
69	Attention/ Concentration	issues due to MH
70	Attention/ Concentration	issues due to DD
71	Attention/ Concentration	varying
72	Attitude	positive/appropriate
73	Attitude	neutral/unable to categorize
74	Attitude	age appropriate
75	Attitude	not age appropriate
76	Attitude	negative/inappropriate
77	Attitude	improved
78	Attitude	declined
79	Cognition	normal/no issues
80	Cognition	neutral/unable to categorize
81	Cognition	issues with concentration
82	Cognition	positive concentration
83	Cognition	issues with fund of knowledge
84	Cognition	positive fund of knowledge
85	Cognition	issues with attention
86	Cognition	positive attention
87	Cognition	general issues
88	Executive Functioning	intact/no issues
89	Executive Functioning	neutral/unable to categorize
90	Executive Functioning	some impairment
91	Executive Functioning	much impairment
92	Executive Functioning	improved since last session
93	Executive Functioning	declined since last session
94	Executive Functioning	age appropriate
95	Executive Functioning	below age level
96	Executive Functioning	impaired (due to DD)

For peer review only

Online supplementary material

97	Executive Functioning	impaired (due to TBI)
98	Executive Functioning	impaired (due to other MH issue)
99	Fund of knowledge	intact/no issues
100	Fund of knowledge	neutral/unable to categorize
101	Fund of knowledge	positive vocabulary
102	Fund of knowledge	issues with vocabulary
103	Fund of knowledge	positive history
104	Fund of knowledge	issues with history
105	Fund of knowledge	average
106	Fund of knowledge	above average
107	Fund of knowledge	below average
108	Fund of knowledge	positive current events
109	Fund of knowledge	issues with current events
110	Fund of knowledge	generally limited/issues
111	Fund of knowledge	declined
112	Fund of knowledge	improved
113	Gait and station	normal/no issues
114	Gait and station	neutral/unable to categorize
115	Gait and station	non-ambulatory
116	Gait and station	issues with gait/station
117	Gait and station	assistive device needed
118	Homicidal	not present
119	Homicidal	neutral
120	Homicidal	present
121	Impulse control	good/no issues
122	Impulse control	neutral/unable to categorize
123	Impulse control	limited/some issues
124	Impulse control	poor/serious issues
125	Impulse control	age appropriate
126	Impulse control	not age appropriate
127	Insight	no issues with insight
128	Insight	issues with insight
129	Insight	neutral/unable to categorize
130	Intelligence	normal/no issues with intelligence
131	Intelligence	issues with intelligence
132	Intelligence	neutral/unable to categorize
133	Judgment	no issues with judgement
134	Judgment	issues with judgement
135	Judgment	neutral/unable to categorize
136	Language	intact
137	Language	neutral/unable to categorize

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

138	Language	repetition intact
139	Language	issues with repetition
140	Language	object naming intact
141	Language	issues with object naming
142	Language	impaired
143	Language	non-verbal/mute
144	Language	minimally verbal
145	Language	issues related to DD
146	Level of consciousness	normal; no issues notes
147	Level of consciousness	neutral; unable to categorize
148	Level of consciousness	sleepy/drowsy
149	Level of consciousness	clouded
150	Level of consciousness	issues verbally
151	Level of consciousness	misc. issues
152	Level of consciousness	impaired due to DD/MI
153	Memory	no issues/memory unimpaired
154	Memory	issues/memory impaired
155	Memory	neutral/unable to categorize
156	Mood	Normal; no issues
157	Mood	neutral; unable to categorize
158	Mood	"anxious, tense"
159	Mood	"depressed, sad, despondent"
160	Mood	"irritable, angry"
161	Mood	improved
162	Mood	declined
163	Mood	"elevated, manic"
164	Mood	labile
165	Mood	misc. issues
166	Mood	suicidal
167	Orientation	no orientation issues
168	Orientation	issues with orientation to person
169	Orientation	issues with orientation to place
170	Orientation	issues with orientation to time/date
171	Orientation	issues with orientation to situation
172	Orientation	neutral/unable to categorize
173	Orientation	general orientation issues

For peer review only

Online supplementary material

174	Psychomotor	normal; no issues
175	Psychomotor	neutral; unable to categorize
176	Psychomotor	retarded; slowed
177	Psychomotor	agitated; tense; restless
178	Psychomotor	improved
179	Psychomotor	declined
180	Psychomotor	constricted
181	Psychomotor	tremors
182	Psychomotor	tics
183	Psychomotor	involuntary movements
184	Psychomotor	misc. issues
185	Psychomotor	catatonic
186	Reasoning	normal; no issues
187	Reasoning	neutral; unable to categorize
188	Reasoning	impaired; poor
189	Reasoning	impaired due to DD/cognitive issues
190	Reasoning	impaired due to MI
191	Reasoning	concrete
192	Reasoning	abstract
193	Reasoning	improved
194	Reasoning	misc. issues
195	Sensorium	normal; no issues
196	Sensorium	neutral; unable to categorize
197	Sensorium	clouded; sedate
198	Sensorium	impaired
199	Sensorium	fluctuating
200	Sensorium	impaired due to medications/drugs/alcohol
201	Sensorium	sleepy; lethargic
202	Sleep	"normal, no issues"
203	Sleep	"neutral, unable to categorize"
204	Sleep	maintenance issues (can't stay asleep)
205	Sleep	onset issues (unable to fall asleep)
206	Sleep	nightmares
207	Sleep	insomnia
208	Sleep	hypersomnia
209	Sleep	misc. issues related to sleep
210	Speech	"normal, no issues"
211	Speech	"neutral, unable to categorize"
212	Speech	issues with rate
213	Speech	issues with tone

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

214	Speech	issues with volume
215	Speech	stutter or lisp
216	Speech	misc. issues
217	Speech	impoverished
218	Speech	improved
219	Speech	unintelligible
220	Speech	"mute, non-verbal"
221	Suicidal	"normal, no issues, not present"
222	Suicidal	"neutral, unable to categorize"
223	Suicidal	suicidal ideation
224	Suicidal	suicidal with intent
225	Suicidal	suicidal with plan
226	Suicidal	suicide attempt
227	Suicidal	history of ideation
228	Suicidal	history of attempt
229	Suicidal	self-injurious
230	Suicidal	history of self-injury
231	Suicidal	suicidal ideation with means
232	Violent Thoughts	"normal, no issues"
233	Violent Thoughts	"neutral, unable to categorize"
234	Violent Thoughts	present
235	Violent Thoughts	"present, with intent"
236	Violent Thoughts	present with plan
237	Violent Thoughts	present with means
238	Violent Thoughts	history of violent thoughts
239	Violent Thoughts	history of violent behaviors
240	Violent Thoughts	violent/aggressive behavior

For peer review only

Online supplementary material

eTable 1: Description of U.S. clinics contributing EHR data to NeuroBlu

<Insert at PDF stage>

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

eTable 2: NeuroBlu dataset race distribution

Race	Number of patients	Percentage
White	276,633	49.1%
Unknown	186,517	33.1%
Black or African American	73,504	13.1%
Other Race	13,179	2.3%
Native Hawaiian or Other Pacific Islander	6,712	1.2%
Asian	3,958	0.7%
American Indian or Alaska Native	2,437	0.4%

eTable 3: NeuroBlu dataset age distribution

Age (years)	Number of patients	Percentage
0-9	45,050	8.0%
10-19	116,741	20.7%
20-29	108,863	19.3%
30-39	94,114	16.7%
40-49	81,556	14.5%
50-59	65,498	11.6%
60-69	30,945	5.5%
70-79	15,339	2.7%
80+	4,828	0.9%

eTable 4: NeuroBlu dataset diagnosis distribution

Diagnosis	Number of patients	Percentage
Dementia & Alzheimer's disease	5,029	0.9%
Substance-use disorders	152,790	28.1%
Schizophrenia	14,592	2.7%
Schizoaffective disorder	15,044	2.8%
Bipolar disorder	69,607	12.8%
Major depressive disorder (MDD)	129,120	23.7%
Anxiety disorders	103,923	19.1%
Attention-deficit hyperactivity disorder (ADHD)	53,744	9.9%

For peer review only

Online supplementary material

eTable 5: Gender distribution by diagnosis

Diagnosis	Number of male patients	Number of female patients	Number with unknown gender	Percentage Male	Percentage Female	Percentage Unknown	Total number
Dementia & Alzheimer's disease	2,035	2,962	32	40.5%	58.9%	0.8%	5,029
Substance-use disorders	96,180	56,328	282	62.9%	36.9%	0.0%	152,790
Schizophrenia	9,163	5,401	28	62.8%	37.0%	0.4%	14,592
Schizoaffective disorder	7,807	7,222	15	51.9%	48.0%	0.3%	15,044
Bipolar disorder	27,877	41,642	88	40.0%	59.8%	0.1%	69,607
Major depressive disorder (MDD)	46,603	82,395	122	36.1%	63.8%	0.0%	129,120
Anxiety disorders	38,255	65,569	99	36.8%	63.1%	0.0%	103,923
Attention-deficit hyperactivity disorder (ADHD)	32,781	20,933	30	61.0%	38.9%	0.1%	53,744

For peer review only

Online supplementary material

eTable 6: ICD-9/ICD-10 code definitions for diagnostic category

Diagnosis	ICD code	ICD type	ICD Description
ADHD	314	ICD-9	Attention deficit disorder without mention of hyperactivity
ADHD	314.01	ICD-9	Attention deficit disorder with hyperactivity
ADHD	F90.0	ICD-10	Attention-deficit hyperactivity disorder, predominantly inattentive type
ADHD	F90.1	ICD-10	Attention-deficit hyperactivity disorder, predominantly hyperactive type
ADHD	F90.2	ICD-10	Attention-deficit hyperactivity disorder, combined type
ADHD	F90.8	ICD-10	Attention-deficit hyperactivity disorder, other type
ADHD	F90.9	ICD-10	Attention-deficit hyperactivity disorder, unspecified type
Anxiety disorders	300.01	ICD-9	Panic disorder without agoraphobia
Anxiety disorders	300.02	ICD-9	Generalized anxiety disorder
Anxiety disorders	300.2	ICD-9	Phobia, unspecified
Anxiety disorders	300.21	ICD-9	Agoraphobia with panic disorder
Anxiety disorders	300.22	ICD-9	Agoraphobia without mention of panic attacks
Anxiety disorders	300.23	ICD-9	Social phobia
Anxiety disorders	300.29	ICD-9	Other isolated or specific phobias
Anxiety disorders	300.3	ICD-9	Obsessive-compulsive disorders
Anxiety disorders	309.81	ICD-9	Posttraumatic stress disorder
Anxiety disorders	F40.00	ICD-10	Agoraphobia, unspecified
Anxiety disorders	F40.01	ICD-10	Agoraphobia with panic disorder
Anxiety disorders	F40.02	ICD-10	Agoraphobia without panic disorder
Anxiety disorders	F40.10	ICD-10	Social phobia, unspecified
Anxiety disorders	F40.11	ICD-10	Social phobia, generalized
Anxiety disorders	F40.210	ICD-10	Arachnophobia
Anxiety disorders	F40.218	ICD-10	Other animal type phobia
Anxiety disorders	F40.220	ICD-10	Fear of thunderstorms
Anxiety disorders	F40.228	ICD-10	Other natural environment type phobia

For peer review only

Online supplementary material

Anxiety disorders	F40.230	ICD-10	Fear of blood
Anxiety disorders	F40.231	ICD-10	Fear of injections and transfusions
Anxiety disorders	F40.232	ICD-10	Fear of other medical care
Anxiety disorders	F40.233	ICD-10	Fear of injury
Anxiety disorders	F40.240	ICD-10	Claustrophobia
Anxiety disorders	F40.241	ICD-10	Acrophobia
Anxiety disorders	F40.242	ICD-10	Fear of bridges
Anxiety disorders	F40.243	ICD-10	Fear of flying
Anxiety disorders	F40.248	ICD-10	Other situational type phobia
Anxiety disorders	F40.290	ICD-10	Androphobia
Anxiety disorders	F40.291	ICD-10	Gynephobia
Anxiety disorders	F40.298	ICD-10	Other specified phobia
Anxiety disorders	F40.8	ICD-10	Other phobic anxiety disorders
Anxiety disorders	F40.9	ICD-10	Phobic anxiety disorder, unspecified
Anxiety disorders	F41.0	ICD-10	Panic disorder [episodic paroxysmal anxiety]
Anxiety disorders	F41.1	ICD-10	Generalized anxiety disorder
Anxiety disorders	F42.2	ICD-10	Mixed obsessional thoughts and acts
Anxiety disorders	F42.3	ICD-10	Hoarding disorder
Anxiety disorders	F42.4	ICD-10	Excoriation (skin-picking) disorder
Anxiety disorders	F42.8	ICD-10	Other obsessive-compulsive disorder
Anxiety disorders	F42.9	ICD-10	Obsessive-compulsive disorder, unspecified
Anxiety disorders	F43.10	ICD-10	Post-traumatic stress disorder, unspecified
Anxiety disorders	F43.11	ICD-10	Post-traumatic stress disorder, acute
Anxiety disorders	F43.12	ICD-10	Post-traumatic stress disorder, chronic
Bipolar Disorder	296	ICD-9	Bipolar I disorder, single manic episode, unspecified

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Bipolar Disorder	296.01	ICD-10	Bipolar I disorder, single manic episode, mild
Bipolar Disorder	296.02	ICD-11	Bipolar I disorder, single manic episode, moderate
Bipolar Disorder	296.03	ICD-12	Bipolar I disorder, single manic episode, severe, without mention of psychotic behavior
Bipolar Disorder	296.04	ICD-13	Bipolar I disorder, single manic episode, specified as with psychotic behavior
Bipolar Disorder	296.05	ICD-14	Bipolar I disorder, single manic episode, in partial or unspecified remission
Bipolar Disorder	296.06	ICD-15	Bipolar I disorder, single manic episode, in full remission
Bipolar Disorder	296.4	ICD-9	Bipolar I disorder, most recent episode (or current) manic, unspecified
Bipolar Disorder	296.41	ICD-9	Bipolar I disorder, most recent episode (or current) manic, mild
Bipolar Disorder	296.42	ICD-9	Bipolar I disorder, most recent episode (or current) manic, moderate
Bipolar Disorder	296.43	ICD-9	Bipolar I disorder, most recent episode (or current) manic, severe, without mention of psychotic behavior
Bipolar Disorder	296.44	ICD-9	Bipolar I disorder, most recent episode (or current) manic, severe, specified as with psychotic behavior
Bipolar Disorder	296.45	ICD-9	Bipolar I disorder, most recent episode (or current) manic, in partial or unspecified remission
Bipolar Disorder	296.46	ICD-9	Bipolar I disorder, most recent episode (or current) manic, in full remission
Bipolar Disorder	296.5	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, unspecified
Bipolar Disorder	296.51	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, mild
Bipolar Disorder	296.52	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, moderate
Bipolar Disorder	296.53	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, severe, without mention of psychotic behavior
Bipolar Disorder	296.54	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, severe, specified as with psychotic behavior
Bipolar Disorder	296.55	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, in partial or unspecified remission
Bipolar Disorder	296.56	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, in full remission
Bipolar Disorder	296.6	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, unspecified
Bipolar Disorder	296.61	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, mild
Bipolar Disorder	296.62	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, moderate
Bipolar Disorder	296.63	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, severe, without mention of psychotic behavior
Bipolar Disorder	296.64	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, severe, specified as with psychotic behavior

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Bipolar Disorder	296.65	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, in partial or unspecified remission
Bipolar Disorder	296.66	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, in full remission
Bipolar Disorder	296.7	ICD-9	Bipolar I disorder, most recent episode (or current) unspecified
Bipolar Disorder	296.8	ICD-9	Bipolar disorder, unspecified
Bipolar Disorder	296.89	ICD-9	Other bipolar disorders
Bipolar Disorder	F31.0	ICD-10	Bipolar disorder, current episode hypomanic
Bipolar Disorder	F31.10	ICD-10	Bipolar disorder, current episode manic without psychotic features, unspecified
Bipolar Disorder	F31.11	ICD-10	Bipolar disorder, current episode manic without psychotic features, mild
Bipolar Disorder	F31.12	ICD-10	Bipolar disorder, current episode manic without psychotic features, moderate
Bipolar Disorder	F31.13	ICD-10	Bipolar disorder, current episode manic without psychotic features, severe
Bipolar Disorder	F31.2	ICD-10	Bipolar disorder, current episode manic severe with psychotic features
Bipolar Disorder	F31.30	ICD-10	Bipolar disorder, current episode depressed, mild or moderate severity, unspecified
Bipolar Disorder	F31.31	ICD-10	Bipolar disorder, current episode depressed, mild
Bipolar Disorder	F31.32	ICD-10	Bipolar disorder, current episode depressed, moderate
Bipolar Disorder	F31.4	ICD-10	Bipolar disorder, current episode depressed, severe, without psychotic features
Bipolar Disorder	F31.5	ICD-10	Bipolar disorder, current episode depressed, severe, with psychotic features
Bipolar Disorder	F31.60	ICD-10	Bipolar disorder, current episode mixed, unspecified
Bipolar Disorder	F31.61	ICD-10	Bipolar disorder, current episode mixed, mild
Bipolar Disorder	F31.62	ICD-10	Bipolar disorder, current episode mixed, moderate
Bipolar Disorder	F31.63	ICD-10	Bipolar disorder, current episode mixed, severe, without psychotic features
Bipolar Disorder	F31.64	ICD-10	Bipolar disorder, current episode mixed, severe, with psychotic features
Bipolar Disorder	F31.70	ICD-10	Bipolar disorder, currently in remission, most recent episode unspecified
Bipolar Disorder	F31.71	ICD-10	Bipolar disorder, in partial remission, most recent episode hypomanic
Bipolar Disorder	F31.72	ICD-10	Bipolar disorder, in full remission, most recent episode hypomanic
Bipolar Disorder	F31.73	ICD-10	Bipolar disorder, in partial remission, most recent episode manic

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Bipolar Disorder	F31.74	ICD-10	Bipolar disorder, in full remission, most recent episode manic
Bipolar Disorder	F31.75	ICD-10	Bipolar disorder, in partial remission, most recent episode depressed
Bipolar Disorder	F31.76	ICD-10	Bipolar disorder, in full remission, most recent episode depressed
Bipolar Disorder	F31.77	ICD-10	Bipolar disorder, in partial remission, most recent episode mixed
Bipolar Disorder	F31.78	ICD-10	Bipolar disorder, in full remission, most recent episode mixed
Bipolar Disorder	F31.81	ICD-10	Bipolar II disorder
Bipolar Disorder	F31.89	ICD-10	Other bipolar disorder
Bipolar Disorder	F31.9	ICD-10	Bipolar disorder, unspecified
Dementia and Alzheimer's Disease	290	ICD-9	Senile dementia, uncomplicated
Dementia and Alzheimer's Disease	290.1	ICD-9	Presenile dementia
Dementia and Alzheimer's Disease	290.2	ICD-9	Senile dementia with delusional or depressive features
Dementia and Alzheimer's Disease	290.3	ICD-9	Senile dementia with delirium
Dementia and Alzheimer's Disease	290.4	ICD-9	Vascular dementia
Dementia and Alzheimer's Disease	290.43	ICD-9	Vascular dementia, with depressed mood
Dementia and Alzheimer's Disease	294.1	ICD-9	Dementia in conditions classified elsewhere
Dementia and Alzheimer's Disease	294.8	ICD-9	Other persistent mental disorders due to conditions classified elsewhere
Dementia and Alzheimer's Disease	331	ICD-9	Alzheimer's disease
Dementia and Alzheimer's Disease	F03.90	ICD-10	Unspecified dementia without behavioral disturbance
Dementia and Alzheimer's Disease	F03.91	ICD-10	Unspecified dementia with behavioral disturbance

For peer review only

Online supplementary material

Dementia and Alzheimer's Disease	G30.1	ICD-10	Alzheimer's disease with late onset
Dementia and Alzheimer's Disease	G30.8	ICD-10	Other Alzheimer's disease
MDD	296.2	ICD-9	Major depressive affective disorder, single episode, unspecified
MDD	296.21	ICD-9	Major depressive affective disorder, single episode, mild
MDD	296.22	ICD-9	Major depressive affective disorder, single episode, moderate
MDD	296.23	ICD-9	Major depressive affective disorder, single episode, severe, without mention of psychotic behavior
MDD	296.24	ICD-9	Major depressive affective disorder, single episode, severe, specified as with psychotic behavior
MDD	296.25	ICD-9	Major depressive affective disorder, single episode, in partial or unspecified remission
MDD	296.26	ICD-9	Major depressive affective disorder, single episode, in full remission
MDD	296.3	ICD-9	Major depressive affective disorder, recurrent episode, unspecified
MDD	296.31	ICD-9	Major depressive affective disorder, recurrent episode, mild
MDD	296.32	ICD-9	Major depressive affective disorder, recurrent episode, moderate
MDD	296.33	ICD-9	Major depressive affective disorder, recurrent episode, severe, without mention of psychotic behavior
MDD	296.34	ICD-9	Major depressive affective disorder, recurrent episode, severe, specified as with psychotic behavior
MDD	296.35	ICD-9	Major depressive affective disorder, recurrent episode, in partial or unspecified remission
MDD	296.36	ICD-9	Major depressive affective disorder, recurrent episode, in full remission
MDD	F32.0	ICD-10	Major depressive disorder, single episode, mild
MDD	F32.1	ICD-10	Major depressive disorder, single episode, moderate
MDD	F32.2	ICD-10	Major depressive disorder, single episode, severe without psychotic features
MDD	F32.3	ICD-10	Major depressive disorder, single episode, severe with psychotic features
MDD	F32.4	ICD-10	Major depressive disorder, single episode, in partial remission
MDD	F32.5	ICD-10	Major depressive disorder, single episode, in full remission
MDD	F32.81	ICD-10	Premenstrual dysphoric disorder
MDD	F32.89	ICD-10	Other specified depressive episodes
MDD	F32.9	ICD-10	Major depressive disorder, single episode, unspecified
MDD	F33.0	ICD-10	Major depressive disorder, recurrent, mild
MDD	F33.1	ICD-10	Major depressive disorder, recurrent, moderate

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

MDD	F33.2	ICD-10	Major depressive disorder, recurrent severe without psychotic features
MDD	F33.3	ICD-10	Major depressive disorder, recurrent, severe with psychotic symptoms
MDD	F33.40	ICD-10	Major depressive disorder, recurrent, in remission, unspecified
MDD	F33.41	ICD-10	Major depressive disorder, recurrent, in partial remission
MDD	F33.42	ICD-10	Major depressive disorder, recurrent, in full remission
MDD	F33.8	ICD-10	Other recurrent depressive disorders
MDD	F33.9	ICD-10	Major depressive disorder, recurrent, unspecified
Schizoaffective disorder	295.7	ICD-9	Schizoaffective disorder, unspecified
Schizoaffective disorder	295.71	ICD-9	Schizoaffective disorder, subchronic
Schizoaffective disorder	295.72	ICD-9	Schizoaffective disorder, chronic
Schizoaffective disorder	295.73	ICD-9	Schizoaffective disorder, subchronic with acute exacerbation
Schizoaffective disorder	295.74	ICD-9	Schizoaffective disorder, chronic with acute exacerbation
Schizoaffective disorder	295.75	ICD-9	Schizoaffective disorder, in remission
Schizoaffective disorder	F25.0	ICD-10	Schizoaffective disorder, bipolar type
Schizoaffective disorder	F25.1	ICD-10	Schizoaffective disorder, depressive type
Schizoaffective disorder	F25.8	ICD-10	Other schizoaffective disorders
Schizoaffective disorder	F25.9	ICD-10	Schizoaffective disorder, unspecified
Schizophrenia	295	ICD-9	Simple type schizophrenia, unspecified
Schizophrenia	295.01	ICD-9	Simple type schizophrenia, subchronic
Schizophrenia	295.02	ICD-9	Simple type schizophrenia, chronic
Schizophrenia	295.03	ICD-9	Simple type schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.04	ICD-9	Simple type schizophrenia, chronic with acute exacerbation
Schizophrenia	295.05	ICD-9	Simple type schizophrenia, in remission
Schizophrenia	295.1	ICD-9	Disorganized type schizophrenia, unspecified
Schizophrenia	295.11	ICD-9	Disorganized type schizophrenia, subchronic
Schizophrenia	295.12	ICD-9	Disorganized type schizophrenia, chronic
Schizophrenia	295.13	ICD-9	Disorganized type schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.14	ICD-9	Disorganized type schizophrenia, chronic with acute exacerbation
Schizophrenia	295.15	ICD-9	Disorganized type schizophrenia, in remission
Schizophrenia	295.2	ICD-9	Catatonic type schizophrenia, unspecified

For peer review only

Online supplementary material

Schizophrenia	295.21	ICD-9	Catatonic type schizophrenia, subchronic
Schizophrenia	295.22	ICD-9	Catatonic type schizophrenia, chronic
Schizophrenia	295.23	ICD-9	Catatonic type schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.24	ICD-9	Catatonic type schizophrenia, chronic with acute exacerbation
Schizophrenia	295.25	ICD-9	Catatonic type schizophrenia, in remission
Schizophrenia	295.3	ICD-9	Paranoid type schizophrenia, unspecified
Schizophrenia	295.31	ICD-9	Paranoid type schizophrenia, subchronic
Schizophrenia	295.32	ICD-9	Paranoid type schizophrenia, chronic
Schizophrenia	295.33	ICD-9	Paranoid type schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.34	ICD-9	Paranoid type schizophrenia, chronic with acute exacerbation
Schizophrenia	295.35	ICD-9	Paranoid type schizophrenia, in remission
Schizophrenia	295.4	ICD-9	Schizophreniform disorder, unspecified
Schizophrenia	295.41	ICD-9	Schizophreniform disorder, subchronic
Schizophrenia	295.42	ICD-9	Schizophreniform disorder, chronic
Schizophrenia	295.43	ICD-9	Schizophreniform disorder, subchronic with acute exacerbation
Schizophrenia	295.44	ICD-9	Schizophreniform disorder, chronic with acute exacerbation
Schizophrenia	295.45	ICD-9	Schizophreniform disorder, in remission
Schizophrenia	295.5	ICD-9	Latent schizophrenia, unspecified
Schizophrenia	295.51	ICD-9	Latent schizophrenia, subchronic
Schizophrenia	295.52	ICD-9	Latent schizophrenia, chronic
Schizophrenia	295.53	ICD-9	Latent schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.54	ICD-9	Latent schizophrenia, chronic with acute exacerbation
Schizophrenia	295.55	ICD-9	Latent schizophrenia, in remission
Schizophrenia	295.6	ICD-9	Schizophrenic disorders, residual type, unspecified
Schizophrenia	295.61	ICD-9	Schizophrenic disorders, residual type, subchronic
Schizophrenia	295.62	ICD-9	Schizophrenic disorders, residual type, chronic
Schizophrenia	295.63	ICD-9	Schizophrenic disorders, residual type, subchronic with acute exacerbation
Schizophrenia	295.64	ICD-9	Schizophrenic disorders, residual type, chronic with acute exacerbation
Schizophrenia	295.65	ICD-9	Schizophrenic disorders, residual type, in remission
Schizophrenia	295.8	ICD-9	Other specified types of schizophrenia, unspecified
Schizophrenia	295.81	ICD-9	Other specified types of schizophrenia, subchronic
Schizophrenia	295.82	ICD-9	Other specified types of schizophrenia, chronic
Schizophrenia	295.83	ICD-9	Other specified types of schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.84	ICD-9	Other specified types of schizophrenia, chronic with acute exacerbation
Schizophrenia	295.85	ICD-9	Other specified types of schizophrenia, in remission

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Schizophrenia	295.9	ICD-9	Unspecified schizophrenia, unspecified
Schizophrenia	295.91	ICD-9	Unspecified schizophrenia, subchronic
Schizophrenia	295.92	ICD-9	Unspecified schizophrenia, chronic
Schizophrenia	295.93	ICD-9	Unspecified schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.94	ICD-9	Unspecified schizophrenia, chronic with acute exacerbation
Schizophrenia	295.95	ICD-9	Unspecified schizophrenia, in remission
Schizophrenia	F20.0	ICD-10	Paranoid schizophrenia
Schizophrenia	F20.1	ICD-10	Disorganized schizophrenia
Schizophrenia	F20.2	ICD-10	Catatonic schizophrenia
Schizophrenia	F20.3	ICD-10	Undifferentiated schizophrenia
Schizophrenia	F20.5	ICD-10	Residual schizophrenia
Schizophrenia	F20.81	ICD-10	Schizophreniform disorder
Schizophrenia	F20.89	ICD-10	Other schizophrenia
Schizophrenia	F20.9	ICD-10	Schizophrenia, unspecified
Substance-use -disorders	291	ICD-9	Alcohol withdrawal delirium
Substance-use -disorders	291.1	ICD-9	Alcohol-induced persisting amnestic disorder
Substance-use -disorders	291.2	ICD-9	Alcohol-induced persisting dementia
Substance-use -disorders	291.3	ICD-9	Alcohol-induced psychotic disorder with hallucinations
Substance-use -disorders	291.4	ICD-9	Idiosyncratic alcohol intoxication
Substance-use -disorders	291.5	ICD-9	Alcohol-induced psychotic disorder with delusions
Substance-use -disorders	291.81	ICD-9	Alcohol withdrawal
Substance-use -disorders	291.82	ICD-9	Alcohol induced sleep disorders
Substance-use -disorders	291.89	ICD-9	Other alcohol-induced mental disorders
Substance-use -disorders	291.9	ICD-9	Unspecified alcohol-induced mental disorders
Substance-use -disorders	292	ICD-9	Drug withdrawal
Substance-use -disorders	292.11	ICD-9	Drug-induced psychotic disorder with delusions
Substance-use -disorders	292.12	ICD-9	Drug-induced psychotic disorder with hallucinations
Substance-use -disorders	292.2	ICD-9	Pathological drug intoxication
Substance-use -disorders	292.81	ICD-9	Drug-induced delirium
Substance-use -disorders	292.82	ICD-9	Drug-induced persisting dementia

For peer review only

Online supplementary material

Substance-use -disorders	292.83	ICD-9	Drug-induced persisting amnestic disorder
Substance-use -disorders	292.84	ICD-9	Drug-induced mood disorder
Substance-use -disorders	292.85	ICD-9	Drug induced sleep disorders
Substance-use -disorders	292.89	ICD-9	Other specified drug-induced mental disorders
Substance-use -disorders	292.9	ICD-9	Unspecified drug-induced mental disorder
Substance-use -disorders	303	ICD-9	Acute alcoholic intoxication in alcoholism, unspecified
Substance-use -disorders	303.01	ICD-9	Acute alcoholic intoxication in alcoholism, continuous
Substance-use -disorders	303.02	ICD-9	Acute alcoholic intoxication in alcoholism, episodic
Substance-use -disorders	303.03	ICD-9	Acute alcoholic intoxication in alcoholism, in remission
Substance-use -disorders	303.9	ICD-9	Other and unspecified alcohol dependence, unspecified
Substance-use -disorders	303.91	ICD-9	Other and unspecified alcohol dependence, continuous
Substance-use -disorders	303.92	ICD-9	Other and unspecified alcohol dependence, episodic
Substance-use -disorders	303.93	ICD-9	Other and unspecified alcohol dependence, in remission
Substance-use -disorders	304	ICD-9	Opioid type dependence, unspecified
Substance-use -disorders	304.01	ICD-9	Opioid type dependence, continuous
Substance-use -disorders	304.02	ICD-9	Opioid type dependence, episodic
Substance-use -disorders	304.03	ICD-9	Opioid type dependence, in remission
Substance-use -disorders	304.1	ICD-9	Sedative, hypnotic or anxiolytic dependence, unspecified
Substance-use -disorders	304.11	ICD-9	Sedative, hypnotic or anxiolytic dependence, continuous
Substance-use -disorders	304.12	ICD-9	Sedative, hypnotic or anxiolytic dependence, episodic
Substance-use -disorders	304.13	ICD-9	Sedative, hypnotic or anxiolytic dependence, in remission
Substance-use -disorders	304.2	ICD-9	Cocaine dependence, unspecified
Substance-use -disorders	304.21	ICD-9	Cocaine dependence, continuous
Substance-use -disorders	304.22	ICD-9	Cocaine dependence, episodic
Substance-use -disorders	304.23	ICD-9	Cocaine dependence, in remission

For peer review only

Online supplementary material

Substance-use -disorders	304.3	ICD-9	Cannabis dependence, unspecified
Substance-use -disorders	304.31	ICD-9	Cannabis dependence, continuous
Substance-use -disorders	304.32	ICD-9	Cannabis dependence, episodic
Substance-use -disorders	304.33	ICD-9	Cannabis dependence, in remission
Substance-use -disorders	304.4	ICD-9	Amphetamine and other psychostimulant dependence, unspecified
Substance-use -disorders	304.41	ICD-9	Amphetamine and other psychostimulant dependence, continuous
Substance-use -disorders	304.42	ICD-9	Amphetamine and other psychostimulant dependence, episodic
Substance-use -disorders	304.43	ICD-9	Amphetamine and other psychostimulant dependence, in remission
Substance-use -disorders	304.5	ICD-9	Hallucinogen dependence, unspecified
Substance-use -disorders	304.51	ICD-9	Hallucinogen dependence, continuous
Substance-use -disorders	304.52	ICD-9	Hallucinogen dependence, episodic
Substance-use -disorders	304.53	ICD-9	Hallucinogen dependence, in remission
Substance-use -disorders	304.6	ICD-9	Other specified drug dependence, unspecified
Substance-use -disorders	304.61	ICD-9	Other specified drug dependence, continuous
Substance-use -disorders	304.62	ICD-9	Other specified drug dependence, episodic
Substance-use -disorders	304.63	ICD-9	Other specified drug dependence, in remission
Substance-use -disorders	304.7	ICD-9	Combinations of opioid type drug with any other drug dependence, unspecified
Substance-use -disorders	304.71	ICD-9	Combinations of opioid type drug with any other drug dependence, continuous
Substance-use -disorders	304.72	ICD-9	Combinations of opioid type drug with any other drug dependence, episodic
Substance-use -disorders	304.73	ICD-9	Combinations of opioid type drug with any other drug dependence, in remission
Substance-use -disorders	304.8	ICD-9	Combinations of drug dependence excluding opioid type drug, unspecified
Substance-use -disorders	304.81	ICD-9	Combinations of drug dependence excluding opioid type drug, continuous
Substance-use -disorders	304.82	ICD-9	Combinations of drug dependence excluding opioid type drug, episodic
Substance-use -disorders	304.83	ICD-9	Combinations of drug dependence excluding opioid type drug, in remission
Substance-use -disorders	304.9	ICD-9	Unspecified drug dependence, unspecified

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	304.91	ICD-9	Unspecified drug dependence, continuous
Substance-use -disorders	304.92	ICD-9	Unspecified drug dependence, episodic
Substance-use -disorders	304.93	ICD-9	Unspecified drug dependence, in remission
Substance-use -disorders	305	ICD-9	Alcohol abuse, unspecified
Substance-use -disorders	305.01	ICD-9	Alcohol abuse, continuous
Substance-use -disorders	305.02	ICD-9	Alcohol abuse, episodic
Substance-use -disorders	305.03	ICD-9	Alcohol abuse, in remission
Substance-use -disorders	305.1	ICD-9	Tobacco use disorder
Substance-use -disorders	305.2	ICD-9	Cannabis abuse, unspecified
Substance-use -disorders	305.21	ICD-9	Cannabis abuse, continuous
Substance-use -disorders	305.22	ICD-9	Cannabis abuse, episodic
Substance-use -disorders	305.23	ICD-9	Cannabis abuse, in remission
Substance-use -disorders	305.3	ICD-9	Hallucinogen abuse, unspecified
Substance-use -disorders	305.31	ICD-9	Hallucinogen abuse, continuous
Substance-use -disorders	305.32	ICD-9	Hallucinogen abuse, episodic
Substance-use -disorders	305.33	ICD-9	Hallucinogen abuse, in remission
Substance-use -disorders	305.4	ICD-9	Sedative, hypnotic or anxiolytic abuse, unspecified
Substance-use -disorders	305.41	ICD-9	Sedative, hypnotic or anxiolytic abuse, continuous
Substance-use -disorders	305.42	ICD-9	Sedative, hypnotic or anxiolytic abuse, episodic
Substance-use -disorders	305.43	ICD-9	Sedative, hypnotic or anxiolytic abuse, in remission
Substance-use -disorders	305.5	ICD-9	Opioid abuse, unspecified
Substance-use -disorders	305.51	ICD-9	Opioid abuse, continuous
Substance-use -disorders	305.52	ICD-9	Opioid abuse, episodic
Substance-use -disorders	305.53	ICD-9	Opioid abuse, in remission
Substance-use -disorders	305.6	ICD-9	Cocaine abuse, unspecified

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	305.61	ICD-9	Cocaine abuse, continuous
Substance-use -disorders	305.62	ICD-9	Cocaine abuse, episodic
Substance-use -disorders	305.63	ICD-9	Cocaine abuse, in remission
Substance-use -disorders	305.7	ICD-9	Amphetamine or related acting sympathomimetic abuse, unspecified
Substance-use -disorders	305.71	ICD-9	Amphetamine or related acting sympathomimetic abuse, continuous
Substance-use -disorders	305.72	ICD-9	Amphetamine or related acting sympathomimetic abuse, episodic
Substance-use -disorders	305.73	ICD-9	Amphetamine or related acting sympathomimetic abuse, in remission
Substance-use -disorders	305.8	ICD-9	Antidepressant type abuse, unspecified
Substance-use -disorders	305.81	ICD-9	Antidepressant type abuse, continuous
Substance-use -disorders	305.82	ICD-9	Antidepressant type abuse, episodic
Substance-use -disorders	305.83	ICD-9	Antidepressant type abuse, in remission
Substance-use -disorders	305.9	ICD-9	Other, mixed, or unspecified drug abuse, unspecified
Substance-use -disorders	305.91	ICD-9	Other, mixed, or unspecified drug abuse, continuous
Substance-use -disorders	305.92	ICD-9	Other, mixed, or unspecified drug abuse, episodic
Substance-use -disorders	305.93	ICD-9	Other, mixed, or unspecified drug abuse, in remission
Substance-use -disorders	F10.10	ICD-10	Alcohol abuse, uncomplicated
Substance-use -disorders	F10.11	ICD-10	Alcohol abuse, in remission
Substance-use -disorders	F10.120	ICD-10	Alcohol abuse with intoxication, uncomplicated
Substance-use -disorders	F10.121	ICD-10	Alcohol abuse with intoxication delirium
Substance-use -disorders	F10.129	ICD-10	Alcohol abuse with intoxication, unspecified
Substance-use -disorders	F10.14	ICD-10	Alcohol abuse with alcohol-induced mood disorder
Substance-use -disorders	F10.150	ICD-10	Alcohol abuse with alcohol-induced psychotic disorder with delusions
Substance-use -disorders	F10.151	ICD-10	Alcohol abuse with alcohol-induced psychotic disorder with hallucinations
Substance-use -disorders	F10.159	ICD-10	Alcohol abuse with alcohol-induced psychotic disorder, unspecified
Substance-use -disorders	F10.180	ICD-10	Alcohol abuse with alcohol-induced anxiety disorder

For peer review only

Online supplementary material

Substance-use -disorders	F10.181	ICD-10	Alcohol abuse with alcohol-induced sexual dysfunction
Substance-use -disorders	F10.182	ICD-10	Alcohol abuse with alcohol-induced sleep disorder
Substance-use -disorders	F10.188	ICD-10	Alcohol abuse with other alcohol-induced disorder
Substance-use -disorders	F10.19	ICD-10	Alcohol abuse with unspecified alcohol-induced disorder
Substance-use -disorders	F10.20	ICD-10	Alcohol dependence, uncomplicated
Substance-use -disorders	F10.21	ICD-10	Alcohol dependence, in remission
Substance-use -disorders	F10.220	ICD-10	Alcohol dependence with intoxication, uncomplicated
Substance-use -disorders	F10.221	ICD-10	Alcohol dependence with intoxication delirium
Substance-use -disorders	F10.229	ICD-10	Alcohol dependence with intoxication, unspecified
Substance-use -disorders	F10.230	ICD-10	Alcohol dependence with withdrawal, uncomplicated
Substance-use -disorders	F10.231	ICD-10	Alcohol dependence with withdrawal delirium
Substance-use -disorders	F10.232	ICD-10	Alcohol dependence with withdrawal with perceptual disturbance
Substance-use -disorders	F10.239	ICD-10	Alcohol dependence with withdrawal, unspecified
Substance-use -disorders	F10.24	ICD-10	Alcohol dependence with alcohol-induced mood disorder
Substance-use -disorders	F10.250	ICD-10	Alcohol dependence with alcohol-induced psychotic disorder with delusions
Substance-use -disorders	F10.251	ICD-10	Alcohol dependence with alcohol-induced psychotic disorder with hallucinations
Substance-use -disorders	F10.259	ICD-10	Alcohol dependence with alcohol-induced psychotic disorder, unspecified
Substance-use -disorders	F10.26	ICD-10	Alcohol dependence with alcohol-induced persisting amnestic disorder
Substance-use -disorders	F10.27	ICD-10	Alcohol dependence with alcohol-induced persisting dementia
Substance-use -disorders	F10.280	ICD-10	Alcohol dependence with alcohol-induced anxiety disorder
Substance-use -disorders	F10.281	ICD-10	Alcohol dependence with alcohol-induced sexual dysfunction
Substance-use -disorders	F10.282	ICD-10	Alcohol dependence with alcohol-induced sleep disorder
Substance-use -disorders	F10.288	ICD-10	Alcohol dependence with other alcohol-induced disorder
Substance-use -disorders	F10.29	ICD-10	Alcohol dependence with unspecified alcohol-induced disorder
Substance-use -disorders	F10.920	ICD-10	Alcohol use, unspecified with intoxication, uncomplicated

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F10.921	ICD-10	Alcohol use, unspecified with intoxication delirium
Substance-use -disorders	F10.929	ICD-10	Alcohol use, unspecified with intoxication, unspecified
Substance-use -disorders	F10.94	ICD-10	Alcohol use, unspecified with alcohol-induced mood disorder
Substance-use -disorders	F10.950	ICD-10	Alcohol use, unspecified with alcohol-induced psychotic disorder with delusions
Substance-use -disorders	F10.951	ICD-10	Alcohol use, unspecified with alcohol-induced psychotic disorder with hallucinations
Substance-use -disorders	F10.959	ICD-10	Alcohol use, unspecified with alcohol-induced psychotic disorder, unspecified
Substance-use -disorders	F10.96	ICD-10	Alcohol use, unspecified with alcohol-induced persisting amnestic disorder
Substance-use -disorders	F10.97	ICD-10	Alcohol use, unspecified with alcohol-induced persisting dementia
Substance-use -disorders	F10.980	ICD-10	Alcohol use, unspecified with alcohol-induced anxiety disorder
Substance-use -disorders	F10.981	ICD-10	Alcohol use, unspecified with alcohol-induced sexual dysfunction
Substance-use -disorders	F10.982	ICD-10	Alcohol use, unspecified with alcohol-induced sleep disorder
Substance-use -disorders	F10.988	ICD-10	Alcohol use, unspecified with other alcohol-induced disorder
Substance-use -disorders	F10.99	ICD-10	Alcohol use, unspecified with unspecified alcohol-induced disorder
Substance-use -disorders	F11.10	ICD-10	Opioid abuse, uncomplicated
Substance-use -disorders	F11.11	ICD-10	Opioid abuse, in remission
Substance-use -disorders	F11.120	ICD-10	Opioid abuse with intoxication, uncomplicated
Substance-use -disorders	F11.121	ICD-10	Opioid abuse with intoxication delirium
Substance-use -disorders	F11.122	ICD-10	Opioid abuse with intoxication with perceptual disturbance
Substance-use -disorders	F11.129	ICD-10	Opioid abuse with intoxication, unspecified
Substance-use -disorders	F11.14	ICD-10	Opioid abuse with opioid-induced mood disorder
Substance-use -disorders	F11.150	ICD-10	Opioid abuse with opioid-induced psychotic disorder with delusions
Substance-use -disorders	F11.151	ICD-10	Opioid abuse with opioid-induced psychotic disorder with hallucinations
Substance-use -disorders	F11.159	ICD-10	Opioid abuse with opioid-induced psychotic disorder, unspecified
Substance-use -disorders	F11.181	ICD-10	Opioid abuse with opioid-induced sexual dysfunction
Substance-use -disorders	F11.182	ICD-10	Opioid abuse with opioid-induced sleep disorder

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F11.188	ICD-10	Opioid abuse with other opioid-induced disorder
Substance-use -disorders	F11.19	ICD-10	Opioid abuse with unspecified opioid-induced disorder
Substance-use -disorders	F11.20	ICD-10	Opioid dependence, uncomplicated
Substance-use -disorders	F11.21	ICD-10	Opioid dependence, in remission
Substance-use -disorders	F11.220	ICD-10	Opioid dependence with intoxication, uncomplicated
Substance-use -disorders	F11.221	ICD-10	Opioid dependence with intoxication delirium
Substance-use -disorders	F11.222	ICD-10	Opioid dependence with intoxication with perceptual disturbance
Substance-use -disorders	F11.229	ICD-10	Opioid dependence with intoxication, unspecified
Substance-use -disorders	F11.23	ICD-10	Opioid dependence with withdrawal
Substance-use -disorders	F11.24	ICD-10	Opioid dependence with opioid-induced mood disorder
Substance-use -disorders	F11.250	ICD-10	Opioid dependence with opioid-induced psychotic disorder with delusions
Substance-use -disorders	F11.251	ICD-10	Opioid dependence with opioid-induced psychotic disorder with hallucinations
Substance-use -disorders	F11.259	ICD-10	Opioid dependence with opioid-induced psychotic disorder, unspecified
Substance-use -disorders	F11.281	ICD-10	Opioid dependence with opioid-induced sexual dysfunction
Substance-use -disorders	F11.282	ICD-10	Opioid dependence with opioid-induced sleep disorder
Substance-use -disorders	F11.288	ICD-10	Opioid dependence with other opioid-induced disorder
Substance-use -disorders	F11.29	ICD-10	Opioid dependence with unspecified opioid-induced disorder
Substance-use -disorders	F11.90	ICD-10	Opioid use, unspecified, uncomplicated
Substance-use -disorders	F11.920	ICD-10	Opioid use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F11.921	ICD-10	Opioid use, unspecified with intoxication delirium
Substance-use -disorders	F11.922	ICD-10	Opioid use, unspecified with intoxication with perceptual disturbance
Substance-use -disorders	F11.929	ICD-10	Opioid use, unspecified with intoxication, unspecified
Substance-use -disorders	F11.93	ICD-10	Opioid use, unspecified with withdrawal
Substance-use -disorders	F11.94	ICD-10	Opioid use, unspecified with opioid-induced mood disorder
Substance-use -disorders	F11.950	ICD-10	Opioid use, unspecified with opioid-induced psychotic disorder with delusions

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F11.951	ICD-10	Opioid use, unspecified with opioid-induced psychotic disorder with hallucinations
Substance-use -disorders	F11.959	ICD-10	Opioid use, unspecified with opioid-induced psychotic disorder, unspecified
Substance-use -disorders	F11.981	ICD-10	Opioid use, unspecified with opioid-induced sexual dysfunction
Substance-use -disorders	F11.982	ICD-10	Opioid use, unspecified with opioid-induced sleep disorder
Substance-use -disorders	F11.988	ICD-10	Opioid use, unspecified with other opioid-induced disorder
Substance-use -disorders	F11.99	ICD-10	Opioid use, unspecified with unspecified opioid-induced disorder
Substance-use -disorders	F12.10	ICD-10	Cannabis abuse, uncomplicated
Substance-use -disorders	F12.11	ICD-10	Cannabis abuse, in remission
Substance-use -disorders	F12.120	ICD-10	Cannabis abuse with intoxication, uncomplicated
Substance-use -disorders	F12.121	ICD-10	Cannabis abuse with intoxication delirium
Substance-use -disorders	F12.122	ICD-10	Cannabis abuse with intoxication with perceptual disturbance
Substance-use -disorders	F12.129	ICD-10	Cannabis abuse with intoxication, unspecified
Substance-use -disorders	F12.150	ICD-10	Cannabis abuse with psychotic disorder with delusions
Substance-use -disorders	F12.151	ICD-10	Cannabis abuse with psychotic disorder with hallucinations
Substance-use -disorders	F12.159	ICD-10	Cannabis abuse with psychotic disorder, unspecified
Substance-use -disorders	F12.180	ICD-10	Cannabis abuse with cannabis-induced anxiety disorder
Substance-use -disorders	F12.188	ICD-10	Cannabis abuse with other cannabis-induced disorder
Substance-use -disorders	F12.19	ICD-10	Cannabis abuse with unspecified cannabis-induced disorder
Substance-use -disorders	F12.20	ICD-10	Cannabis dependence, uncomplicated
Substance-use -disorders	F12.21	ICD-10	Cannabis dependence, in remission
Substance-use -disorders	F12.220	ICD-10	Cannabis dependence with intoxication, uncomplicated
Substance-use -disorders	F12.221	ICD-10	Cannabis dependence with intoxication delirium
Substance-use -disorders	F12.222	ICD-10	Cannabis dependence with intoxication with perceptual disturbance
Substance-use -disorders	F12.229	ICD-10	Cannabis dependence with intoxication, unspecified
Substance-use -disorders	F12.250	ICD-10	Cannabis dependence with psychotic disorder with delusions

For peer review only

Online supplementary material

Substance-use -disorders	F12.251	ICD-10	Cannabis dependence with psychotic disorder with hallucinations
Substance-use -disorders	F12.259	ICD-10	Cannabis dependence with psychotic disorder, unspecified
Substance-use -disorders	F12.280	ICD-10	Cannabis dependence with cannabis-induced anxiety disorder
Substance-use -disorders	F12.288	ICD-10	Cannabis dependence with other cannabis-induced disorder
Substance-use -disorders	F12.29	ICD-10	Cannabis dependence with unspecified cannabis-induced disorder
Substance-use -disorders	F12.90	ICD-10	Cannabis use, unspecified, uncomplicated
Substance-use -disorders	F12.920	ICD-10	Cannabis use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F12.921	ICD-10	Cannabis use, unspecified with intoxication delirium
Substance-use -disorders	F12.922	ICD-10	Cannabis use, unspecified with intoxication with perceptual disturbance
Substance-use -disorders	F12.929	ICD-10	Cannabis use, unspecified with intoxication, unspecified
Substance-use -disorders	F12.950	ICD-10	Cannabis use, unspecified with psychotic disorder with delusions
Substance-use -disorders	F12.951	ICD-10	Cannabis use, unspecified with psychotic disorder with hallucinations
Substance-use -disorders	F12.959	ICD-10	Cannabis use, unspecified with psychotic disorder, unspecified
Substance-use -disorders	F12.980	ICD-10	Cannabis use, unspecified with anxiety disorder
Substance-use -disorders	F12.988	ICD-10	Cannabis use, unspecified with other cannabis-induced disorder
Substance-use -disorders	F12.99	ICD-10	Cannabis use, unspecified with unspecified cannabis-induced disorder
Substance-use -disorders	F13.10	ICD-10	Sedative, hypnotic or anxiolytic abuse, uncomplicated
Substance-use -disorders	F13.11	ICD-10	Sedative, hypnotic or anxiolytic abuse, in remission
Substance-use -disorders	F13.120	ICD-10	Sedative, hypnotic or anxiolytic abuse with intoxication, uncomplicated
Substance-use -disorders	F13.121	ICD-10	Sedative, hypnotic or anxiolytic abuse with intoxication delirium
Substance-use -disorders	F13.129	ICD-10	Sedative, hypnotic or anxiolytic abuse with intoxication, unspecified
Substance-use -disorders	F13.14	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced mood disorder
Substance-use -disorders	F13.150	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced psychotic disorder with delusions
Substance-use -disorders	F13.151	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced psychotic disorder with hallucinations

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F13.159	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced psychotic disorder, unspecified
Substance-use -disorders	F13.180	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced anxiety disorder
Substance-use -disorders	F13.181	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced sexual dysfunction
Substance-use -disorders	F13.182	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced sleep disorder
Substance-use -disorders	F13.188	ICD-10	Sedative, hypnotic or anxiolytic abuse with other sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F13.19	ICD-10	Sedative, hypnotic or anxiolytic abuse with unspecified sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F13.20	ICD-10	Sedative, hypnotic or anxiolytic dependence, uncomplicated
Substance-use -disorders	F13.21	ICD-10	Sedative, hypnotic or anxiolytic dependence, in remission
Substance-use -disorders	F13.220	ICD-10	Sedative, hypnotic or anxiolytic dependence with intoxication, uncomplicated
Substance-use -disorders	F13.221	ICD-10	Sedative, hypnotic or anxiolytic dependence with intoxication delirium
Substance-use -disorders	F13.229	ICD-10	Sedative, hypnotic or anxiolytic dependence with intoxication, unspecified
Substance-use -disorders	F13.230	ICD-10	Sedative, hypnotic or anxiolytic dependence with withdrawal, uncomplicated
Substance-use -disorders	F13.231	ICD-10	Sedative, hypnotic or anxiolytic dependence with withdrawal delirium
Substance-use -disorders	F13.232	ICD-10	Sedative, hypnotic or anxiolytic dependence with withdrawal with perceptual disturbance
Substance-use -disorders	F13.239	ICD-10	Sedative, hypnotic or anxiolytic dependence with withdrawal, unspecified
Substance-use -disorders	F13.24	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced mood disorder
Substance-use -disorders	F13.250	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced psychotic disorder with delusions
Substance-use -disorders	F13.251	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced psychotic disorder with hallucinations
Substance-use -disorders	F13.259	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced psychotic disorder, unspecified
Substance-use -disorders	F13.26	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced persisting amnestic disorder
Substance-use -disorders	F13.27	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced persisting dementia
Substance-use -disorders	F13.280	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced anxiety disorder

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F13.281	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced sexual dysfunction
Substance-use -disorders	F13.282	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced sleep disorder
Substance-use -disorders	F13.288	ICD-10	Sedative, hypnotic or anxiolytic dependence with other sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F13.29	ICD-10	Sedative, hypnotic or anxiolytic dependence with unspecified sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F13.90	ICD-10	Sedative, hypnotic, or anxiolytic use, unspecified, uncomplicated
Substance-use -disorders	F13.920	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F13.921	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with intoxication delirium
Substance-use -disorders	F13.929	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with intoxication, unspecified
Substance-use -disorders	F13.930	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with withdrawal, uncomplicated
Substance-use -disorders	F13.931	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with withdrawal delirium
Substance-use -disorders	F13.932	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with withdrawal with perceptual disturbances
Substance-use -disorders	F13.939	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with withdrawal, unspecified
Substance-use -disorders	F13.94	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced mood disorder
Substance-use -disorders	F13.950	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced psychotic disorder with delusions
Substance-use -disorders	F13.951	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced psychotic disorder with hallucinations
Substance-use -disorders	F13.959	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced psychotic disorder, unspecified
Substance-use -disorders	F13.96	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced persisting amnestic disorder
Substance-use -disorders	F13.97	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced persisting dementia
Substance-use -disorders	F13.980	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced anxiety disorder
Substance-use -disorders	F13.981	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced sexual dysfunction
Substance-use -disorders	F13.982	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced sleep disorder

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F13.988	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with other sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F13.99	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with unspecified sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F14.10	ICD-10	Cocaine abuse, uncomplicated
Substance-use -disorders	F14.11	ICD-10	Cocaine abuse, in remission
Substance-use -disorders	F14.120	ICD-10	Cocaine abuse with intoxication, uncomplicated
Substance-use -disorders	F14.121	ICD-10	Cocaine abuse with intoxication with delirium
Substance-use -disorders	F14.122	ICD-10	Cocaine abuse with intoxication with perceptual disturbance
Substance-use -disorders	F14.129	ICD-10	Cocaine abuse with intoxication, unspecified
Substance-use -disorders	F14.14	ICD-10	Cocaine abuse with cocaine-induced mood disorder
Substance-use -disorders	F14.150	ICD-10	Cocaine abuse with cocaine-induced psychotic disorder with delusions
Substance-use -disorders	F14.151	ICD-10	Cocaine abuse with cocaine-induced psychotic disorder with hallucinations
Substance-use -disorders	F14.159	ICD-10	Cocaine abuse with cocaine-induced psychotic disorder, unspecified
Substance-use -disorders	F14.180	ICD-10	Cocaine abuse with cocaine-induced anxiety disorder
Substance-use -disorders	F14.181	ICD-10	Cocaine abuse with cocaine-induced sexual dysfunction
Substance-use -disorders	F14.182	ICD-10	Cocaine abuse with cocaine-induced sleep disorder
Substance-use -disorders	F14.188	ICD-10	Cocaine abuse with other cocaine-induced disorder
Substance-use -disorders	F14.19	ICD-10	Cocaine abuse with unspecified cocaine-induced disorder
Substance-use -disorders	F14.20	ICD-10	Cocaine dependence, uncomplicated
Substance-use -disorders	F14.21	ICD-10	Cocaine dependence, in remission
Substance-use -disorders	F14.220	ICD-10	Cocaine dependence with intoxication, uncomplicated
Substance-use -disorders	F14.221	ICD-10	Cocaine dependence with intoxication delirium
Substance-use -disorders	F14.222	ICD-10	Cocaine dependence with intoxication with perceptual disturbance
Substance-use -disorders	F14.229	ICD-10	Cocaine dependence with intoxication, unspecified
Substance-use -disorders	F14.23	ICD-10	Cocaine dependence with withdrawal
Substance-use -disorders	F14.24	ICD-10	Cocaine dependence with cocaine-induced mood disorder

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F14.250	ICD-10	Cocaine dependence with cocaine-induced psychotic disorder with delusions
Substance-use -disorders	F14.251	ICD-10	Cocaine dependence with cocaine-induced psychotic disorder with hallucinations
Substance-use -disorders	F14.259	ICD-10	Cocaine dependence with cocaine-induced psychotic disorder, unspecified
Substance-use -disorders	F14.280	ICD-10	Cocaine dependence with cocaine-induced anxiety disorder
Substance-use -disorders	F14.281	ICD-10	Cocaine dependence with cocaine-induced sexual dysfunction
Substance-use -disorders	F14.282	ICD-10	Cocaine dependence with cocaine-induced sleep disorder
Substance-use -disorders	F14.288	ICD-10	Cocaine dependence with other cocaine-induced disorder
Substance-use -disorders	F14.29	ICD-10	Cocaine dependence with unspecified cocaine-induced disorder
Substance-use -disorders	F14.90	ICD-10	Cocaine use, unspecified, uncomplicated
Substance-use -disorders	F14.920	ICD-10	Cocaine use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F14.921	ICD-10	Cocaine use, unspecified with intoxication delirium
Substance-use -disorders	F14.922	ICD-10	Cocaine use, unspecified with intoxication with perceptual disturbance
Substance-use -disorders	F14.929	ICD-10	Cocaine use, unspecified with intoxication, unspecified
Substance-use -disorders	F14.94	ICD-10	Cocaine use, unspecified with cocaine-induced mood disorder
Substance-use -disorders	F14.950	ICD-10	Cocaine use, unspecified with cocaine-induced psychotic disorder with delusions
Substance-use -disorders	F14.951	ICD-10	Cocaine use, unspecified with cocaine-induced psychotic disorder with hallucinations
Substance-use -disorders	F14.959	ICD-10	Cocaine use, unspecified with cocaine-induced psychotic disorder, unspecified
Substance-use -disorders	F14.980	ICD-10	Cocaine use, unspecified with cocaine-induced anxiety disorder
Substance-use -disorders	F14.981	ICD-10	Cocaine use, unspecified with cocaine-induced sexual dysfunction
Substance-use -disorders	F14.982	ICD-10	Cocaine use, unspecified with cocaine-induced sleep disorder
Substance-use -disorders	F14.988	ICD-10	Cocaine use, unspecified with other cocaine-induced disorder
Substance-use -disorders	F14.99	ICD-10	Cocaine use, unspecified with unspecified cocaine-induced disorder
Substance-use -disorders	F15.10	ICD-10	Other stimulant abuse, uncomplicated
Substance-use -disorders	F15.11	ICD-10	Other stimulant abuse, in remission
Substance-use -disorders	F15.120	ICD-10	Other stimulant abuse with intoxication, uncomplicated

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F15.121	ICD-10	Other stimulant abuse with intoxication delirium
Substance-use -disorders	F15.122	ICD-10	Other stimulant abuse with intoxication with perceptual disturbance
Substance-use -disorders	F15.129	ICD-10	Other stimulant abuse with intoxication, unspecified
Substance-use -disorders	F15.14	ICD-10	Other stimulant abuse with stimulant-induced mood disorder
Substance-use -disorders	F15.150	ICD-10	Other stimulant abuse with stimulant-induced psychotic disorder with delusions
Substance-use -disorders	F15.151	ICD-10	Other stimulant abuse with stimulant-induced psychotic disorder with hallucinations
Substance-use -disorders	F15.159	ICD-10	Other stimulant abuse with stimulant-induced psychotic disorder, unspecified
Substance-use -disorders	F15.180	ICD-10	Other stimulant abuse with stimulant-induced anxiety disorder
Substance-use -disorders	F15.181	ICD-10	Other stimulant abuse with stimulant-induced sexual dysfunction
Substance-use -disorders	F15.182	ICD-10	Other stimulant abuse with stimulant-induced sleep disorder
Substance-use -disorders	F15.188	ICD-10	Other stimulant abuse with other stimulant-induced disorder
Substance-use -disorders	F15.19	ICD-10	Other stimulant abuse with unspecified stimulant-induced disorder
Substance-use -disorders	F15.20	ICD-10	Other stimulant dependence, uncomplicated
Substance-use -disorders	F15.21	ICD-10	Other stimulant dependence, in remission
Substance-use -disorders	F15.220	ICD-10	Other stimulant dependence with intoxication, uncomplicated
Substance-use -disorders	F15.221	ICD-10	Other stimulant dependence with intoxication delirium
Substance-use -disorders	F15.222	ICD-10	Other stimulant dependence with intoxication with perceptual disturbance
Substance-use -disorders	F15.229	ICD-10	Other stimulant dependence with intoxication, unspecified
Substance-use -disorders	F15.23	ICD-10	Other stimulant dependence with withdrawal
Substance-use -disorders	F15.24	ICD-10	Other stimulant dependence with stimulant-induced mood disorder
Substance-use -disorders	F15.250	ICD-10	Other stimulant dependence with stimulant-induced psychotic disorder with delusions
Substance-use -disorders	F15.251	ICD-10	Other stimulant dependence with stimulant-induced psychotic disorder with hallucinations
Substance-use -disorders	F15.259	ICD-10	Other stimulant dependence with stimulant-induced psychotic disorder, unspecified
Substance-use -disorders	F15.280	ICD-10	Other stimulant dependence with stimulant-induced anxiety disorder
Substance-use -disorders	F15.281	ICD-10	Other stimulant dependence with stimulant-induced sexual dysfunction

For peer review only

Online supplementary material

Substance-use -disorders	F15.282	ICD-10	Other stimulant dependence with stimulant-induced sleep disorder
Substance-use -disorders	F15.288	ICD-10	Other stimulant dependence with other stimulant-induced disorder
Substance-use -disorders	F15.29	ICD-10	Other stimulant dependence with unspecified stimulant-induced disorder
Substance-use -disorders	F15.90	ICD-10	Other stimulant use, unspecified, uncomplicated
Substance-use -disorders	F15.920	ICD-10	Other stimulant use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F15.921	ICD-10	Other stimulant use, unspecified with intoxication delirium
Substance-use -disorders	F15.922	ICD-10	Other stimulant use, unspecified with intoxication with perceptual disturbance
Substance-use -disorders	F15.929	ICD-10	Other stimulant use, unspecified with intoxication, unspecified
Substance-use -disorders	F15.93	ICD-10	Other stimulant use, unspecified with withdrawal
Substance-use -disorders	F15.94	ICD-10	Other stimulant use, unspecified with stimulant-induced mood disorder
Substance-use -disorders	F15.950	ICD-10	Other stimulant use, unspecified with stimulant-induced psychotic disorder with delusions
Substance-use -disorders	F15.951	ICD-10	Other stimulant use, unspecified with stimulant-induced psychotic disorder with hallucinations
Substance-use -disorders	F15.959	ICD-10	Other stimulant use, unspecified with stimulant-induced psychotic disorder, unspecified
Substance-use -disorders	F15.980	ICD-10	Other stimulant use, unspecified with stimulant-induced anxiety disorder
Substance-use -disorders	F15.981	ICD-10	Other stimulant use, unspecified with stimulant-induced sexual dysfunction
Substance-use -disorders	F15.982	ICD-10	Other stimulant use, unspecified with stimulant-induced sleep disorder
Substance-use -disorders	F15.988	ICD-10	Other stimulant use, unspecified with other stimulant-induced disorder
Substance-use -disorders	F15.99	ICD-10	Other stimulant use, unspecified with unspecified stimulant-induced disorder
Substance-use -disorders	F16.10	ICD-10	Hallucinogen abuse, uncomplicated
Substance-use -disorders	F16.11	ICD-10	Hallucinogen abuse, in remission
Substance-use -disorders	F16.120	ICD-10	Hallucinogen abuse with intoxication, uncomplicated
Substance-use -disorders	F16.121	ICD-10	Hallucinogen abuse with intoxication with delirium
Substance-use -disorders	F16.122	ICD-10	Hallucinogen abuse with intoxication with perceptual disturbance
Substance-use -disorders	F16.129	ICD-10	Hallucinogen abuse with intoxication, unspecified
Substance-use -disorders	F16.14	ICD-10	Hallucinogen abuse with hallucinogen-induced mood disorder

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F16.150	ICD-10	Hallucinogen abuse with hallucinogen-induced psychotic disorder with delusions
Substance-use -disorders	F16.151	ICD-10	Hallucinogen abuse with hallucinogen-induced psychotic disorder with hallucinations
Substance-use -disorders	F16.159	ICD-10	Hallucinogen abuse with hallucinogen-induced psychotic disorder, unspecified
Substance-use -disorders	F16.180	ICD-10	Hallucinogen abuse with hallucinogen-induced anxiety disorder
Substance-use -disorders	F16.183	ICD-10	Hallucinogen abuse with hallucinogen persisting perception disorder (flashbacks)
Substance-use -disorders	F16.188	ICD-10	Hallucinogen abuse with other hallucinogen-induced disorder
Substance-use -disorders	F16.19	ICD-10	Hallucinogen abuse with unspecified hallucinogen-induced disorder
Substance-use -disorders	F16.20	ICD-10	Hallucinogen dependence, uncomplicated
Substance-use -disorders	F16.21	ICD-10	Hallucinogen dependence, in remission
Substance-use -disorders	F16.220	ICD-10	Hallucinogen dependence with intoxication, uncomplicated
Substance-use -disorders	F16.221	ICD-10	Hallucinogen dependence with intoxication with delirium
Substance-use -disorders	F16.229	ICD-10	Hallucinogen dependence with intoxication, unspecified
Substance-use -disorders	F16.24	ICD-10	Hallucinogen dependence with hallucinogen-induced mood disorder
Substance-use -disorders	F16.250	ICD-10	Hallucinogen dependence with hallucinogen-induced psychotic disorder with delusions
Substance-use -disorders	F16.251	ICD-10	Hallucinogen dependence with hallucinogen-induced psychotic disorder with hallucinations
Substance-use -disorders	F16.259	ICD-10	Hallucinogen dependence with hallucinogen-induced psychotic disorder, unspecified
Substance-use -disorders	F16.280	ICD-10	Hallucinogen dependence with hallucinogen-induced anxiety disorder
Substance-use -disorders	F16.283	ICD-10	Hallucinogen dependence with hallucinogen persisting perception disorder (flashbacks)
Substance-use -disorders	F16.288	ICD-10	Hallucinogen dependence with other hallucinogen-induced disorder
Substance-use -disorders	F16.29	ICD-10	Hallucinogen dependence with unspecified hallucinogen-induced disorder
Substance-use -disorders	F16.90	ICD-10	Hallucinogen use, unspecified, uncomplicated
Substance-use -disorders	F16.920	ICD-10	Hallucinogen use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F16.921	ICD-10	Hallucinogen use, unspecified with intoxication with delirium
Substance-use -disorders	F16.929	ICD-10	Hallucinogen use, unspecified with intoxication, unspecified
Substance-use -disorders	F16.94	ICD-10	Hallucinogen use, unspecified with hallucinogen-induced mood disorder

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F16.950	ICD-10	Hallucinogen use, unspecified with hallucinogen-induced psychotic disorder with delusions
Substance-use -disorders	F16.951	ICD-10	Hallucinogen use, unspecified with hallucinogen-induced psychotic disorder with hallucinations
Substance-use -disorders	F16.959	ICD-10	Hallucinogen use, unspecified with hallucinogen-induced psychotic disorder, unspecified
Substance-use -disorders	F16.980	ICD-10	Hallucinogen use, unspecified with hallucinogen-induced anxiety disorder
Substance-use -disorders	F16.983	ICD-10	Hallucinogen use, unspecified with hallucinogen persisting perception disorder (flashbacks)
Substance-use -disorders	F16.988	ICD-10	Hallucinogen use, unspecified with other hallucinogen-induced disorder
Substance-use -disorders	F16.99	ICD-10	Hallucinogen use, unspecified with unspecified hallucinogen-induced disorder
Substance-use -disorders	F17.200	ICD-10	Nicotine dependence, unspecified, uncomplicated
Substance-use -disorders	F17.201	ICD-10	Nicotine dependence, unspecified, in remission
Substance-use -disorders	F17.203	ICD-10	Nicotine dependence unspecified, with withdrawal
Substance-use -disorders	F17.208	ICD-10	Nicotine dependence, unspecified, with other nicotine-induced disorders
Substance-use -disorders	F17.209	ICD-10	Nicotine dependence, unspecified, with unspecified nicotine-induced disorders
Substance-use -disorders	F17.210	ICD-10	Nicotine dependence, cigarettes, uncomplicated
Substance-use -disorders	F17.211	ICD-10	Nicotine dependence, cigarettes, in remission
Substance-use -disorders	F17.213	ICD-10	Nicotine dependence, cigarettes, with withdrawal
Substance-use -disorders	F17.218	ICD-10	Nicotine dependence, cigarettes, with other nicotine-induced disorders
Substance-use -disorders	F17.219	ICD-10	Nicotine dependence, cigarettes, with unspecified nicotine-induced disorders
Substance-use -disorders	F17.220	ICD-10	Nicotine dependence, chewing tobacco, uncomplicated
Substance-use -disorders	F17.221	ICD-10	Nicotine dependence, chewing tobacco, in remission
Substance-use -disorders	F17.223	ICD-10	Nicotine dependence, chewing tobacco, with withdrawal
Substance-use -disorders	F17.228	ICD-10	Nicotine dependence, chewing tobacco, with other nicotine-induced disorders
Substance-use -disorders	F17.229	ICD-10	Nicotine dependence, chewing tobacco, with unspecified nicotine-induced disorders
Substance-use -disorders	F17.290	ICD-10	Nicotine dependence, other tobacco product, uncomplicated
Substance-use -disorders	F17.291	ICD-10	Nicotine dependence, other tobacco product, in remission
Substance-use -disorders	F17.293	ICD-10	Nicotine dependence, other tobacco product, with withdrawal

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F17.298	ICD-10	Nicotine dependence, other tobacco product, with other nicotine-induced disorders
Substance-use -disorders	F17.299	ICD-10	Nicotine dependence, other tobacco product, with unspecified nicotine-induced disorders
Substance-use -disorders	F18.10	ICD-10	Inhalant abuse, uncomplicated
Substance-use -disorders	F18.11	ICD-10	Inhalant abuse, in remission
Substance-use -disorders	F18.120	ICD-10	Inhalant abuse with intoxication, uncomplicated
Substance-use -disorders	F18.121	ICD-10	Inhalant abuse with intoxication delirium
Substance-use -disorders	F18.129	ICD-10	Inhalant abuse with intoxication, unspecified
Substance-use -disorders	F18.14	ICD-10	Inhalant abuse with inhalant-induced mood disorder
Substance-use -disorders	F18.150	ICD-10	Inhalant abuse with inhalant-induced psychotic disorder with delusions
Substance-use -disorders	F18.151	ICD-10	Inhalant abuse with inhalant-induced psychotic disorder with hallucinations
Substance-use -disorders	F18.159	ICD-10	Inhalant abuse with inhalant-induced psychotic disorder, unspecified
Substance-use -disorders	F18.17	ICD-10	Inhalant abuse with inhalant-induced dementia
Substance-use -disorders	F18.180	ICD-10	Inhalant abuse with inhalant-induced anxiety disorder
Substance-use -disorders	F18.188	ICD-10	Inhalant abuse with other inhalant-induced disorder
Substance-use -disorders	F18.19	ICD-10	Inhalant abuse with unspecified inhalant-induced disorder
Substance-use -disorders	F18.20	ICD-10	Inhalant dependence, uncomplicated
Substance-use -disorders	F18.21	ICD-10	Inhalant dependence, in remission
Substance-use -disorders	F18.220	ICD-10	Inhalant dependence with intoxication, uncomplicated
Substance-use -disorders	F18.221	ICD-10	Inhalant dependence with intoxication delirium
Substance-use -disorders	F18.229	ICD-10	Inhalant dependence with intoxication, unspecified
Substance-use -disorders	F18.24	ICD-10	Inhalant dependence with inhalant-induced mood disorder
Substance-use -disorders	F18.250	ICD-10	Inhalant dependence with inhalant-induced psychotic disorder with delusions
Substance-use -disorders	F18.251	ICD-10	Inhalant dependence with inhalant-induced psychotic disorder with hallucinations
Substance-use -disorders	F18.259	ICD-10	Inhalant dependence with inhalant-induced psychotic disorder, unspecified
Substance-use -disorders	F18.27	ICD-10	Inhalant dependence with inhalant-induced dementia

For peer review only

Online supplementary material

Substance-use -disorders	F18.280	ICD-10	Inhalant dependence with inhalant-induced anxiety disorder
Substance-use -disorders	F18.288	ICD-10	Inhalant dependence with other inhalant-induced disorder
Substance-use -disorders	F18.29	ICD-10	Inhalant dependence with unspecified inhalant-induced disorder
Substance-use -disorders	F18.90	ICD-10	Inhalant use, unspecified, uncomplicated
Substance-use -disorders	F18.920	ICD-10	Inhalant use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F18.921	ICD-10	Inhalant use, unspecified with intoxication with delirium
Substance-use -disorders	F18.929	ICD-10	Inhalant use, unspecified with intoxication, unspecified
Substance-use -disorders	F18.94	ICD-10	Inhalant use, unspecified with inhalant-induced mood disorder
Substance-use -disorders	F18.950	ICD-10	Inhalant use, unspecified with inhalant-induced psychotic disorder with delusions
Substance-use -disorders	F18.951	ICD-10	Inhalant use, unspecified with inhalant-induced psychotic disorder with hallucinations
Substance-use -disorders	F18.959	ICD-10	Inhalant use, unspecified with inhalant-induced psychotic disorder, unspecified
Substance-use -disorders	F18.97	ICD-10	Inhalant use, unspecified with inhalant-induced persisting dementia
Substance-use -disorders	F18.980	ICD-10	Inhalant use, unspecified with inhalant-induced anxiety disorder
Substance-use -disorders	F18.988	ICD-10	Inhalant use, unspecified with other inhalant-induced disorder
Substance-use -disorders	F18.99	ICD-10	Inhalant use, unspecified with unspecified inhalant-induced disorder
Substance-use -disorders	F19.10	ICD-10	Other psychoactive substance abuse, uncomplicated
Substance-use -disorders	F19.11	ICD-10	Other psychoactive substance abuse, in remission
Substance-use -disorders	F19.120	ICD-10	Other psychoactive substance abuse with intoxication, uncomplicated
Substance-use -disorders	F19.121	ICD-10	Other psychoactive substance abuse with intoxication delirium
Substance-use -disorders	F19.122	ICD-10	Other psychoactive substance abuse with intoxication with perceptual disturbances
Substance-use -disorders	F19.129	ICD-10	Other psychoactive substance abuse with intoxication, unspecified
Substance-use -disorders	F19.14	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced mood disorder
Substance-use -disorders	F19.150	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced psychotic disorder with delusions
Substance-use -disorders	F19.151	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced psychotic disorder with hallucinations
Substance-use -disorders	F19.159	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced psychotic disorder, unspecified

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F19.16	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced persisting amnestic disorder
Substance-use -disorders	F19.17	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced persisting dementia
Substance-use -disorders	F19.180	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced anxiety disorder
Substance-use -disorders	F19.181	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced sexual dysfunction
Substance-use -disorders	F19.182	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced sleep disorder
Substance-use -disorders	F19.188	ICD-10	Other psychoactive substance abuse with other psychoactive substance-induced disorder
Substance-use -disorders	F19.19	ICD-10	Other psychoactive substance abuse with unspecified psychoactive substance-induced disorder
Substance-use -disorders	F19.20	ICD-10	Other psychoactive substance dependence, uncomplicated
Substance-use -disorders	F19.21	ICD-10	Other psychoactive substance dependence, in remission
Substance-use -disorders	F19.220	ICD-10	Other psychoactive substance dependence with intoxication, uncomplicated
Substance-use -disorders	F19.221	ICD-10	Other psychoactive substance dependence with intoxication delirium
Substance-use -disorders	F19.222	ICD-10	Other psychoactive substance dependence with intoxication with perceptual disturbance
Substance-use -disorders	F19.229	ICD-10	Other psychoactive substance dependence with intoxication, unspecified
Substance-use -disorders	F19.230	ICD-10	Other psychoactive substance dependence with withdrawal, uncomplicated
Substance-use -disorders	F19.231	ICD-10	Other psychoactive substance dependence with withdrawal delirium
Substance-use -disorders	F19.232	ICD-10	Other psychoactive substance dependence with withdrawal with perceptual disturbance
Substance-use -disorders	F19.239	ICD-10	Other psychoactive substance dependence with withdrawal, unspecified
Substance-use -disorders	F19.24	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced mood disorder
Substance-use -disorders	F19.250	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced psychotic disorder with delusions
Substance-use -disorders	F19.251	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced psychotic disorder with hallucinations
Substance-use -disorders	F19.259	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced psychotic disorder, unspecified
Substance-use -disorders	F19.26	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced persisting amnestic disorder
Substance-use -disorders	F19.27	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced persisting dementia

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F19.280	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced anxiety disorder
Substance-use -disorders	F19.281	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced sexual dysfunction
Substance-use -disorders	F19.282	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced sleep disorder
Substance-use -disorders	F19.288	ICD-10	Other psychoactive substance dependence with other psychoactive substance-induced disorder
Substance-use -disorders	F19.29	ICD-10	Other psychoactive substance dependence with unspecified psychoactive substance-induced disorder
Substance-use -disorders	F19.90	ICD-10	Other psychoactive substance use, unspecified, uncomplicated
Substance-use -disorders	F19.920	ICD-10	Other psychoactive substance use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F19.921	ICD-10	Other psychoactive substance use, unspecified with intoxication with delirium
Substance-use -disorders	F19.922	ICD-10	Other psychoactive substance use, unspecified with intoxication with perceptual disturbance
Substance-use -disorders	F19.929	ICD-10	Other psychoactive substance use, unspecified with intoxication, unspecified
Substance-use -disorders	F19.930	ICD-10	Other psychoactive substance use, unspecified with withdrawal, uncomplicated
Substance-use -disorders	F19.931	ICD-10	Other psychoactive substance use, unspecified with withdrawal delirium
Substance-use -disorders	F19.932	ICD-10	Other psychoactive substance use, unspecified with withdrawal with perceptual disturbance
Substance-use -disorders	F19.939	ICD-10	Other psychoactive substance use, unspecified with withdrawal, unspecified
Substance-use -disorders	F19.94	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced mood disorder
Substance-use -disorders	F19.950	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced psychotic disorder with delusions
Substance-use -disorders	F19.951	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced psychotic disorder with hallucinations
Substance-use -disorders	F19.959	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced psychotic disorder, unspecified
Substance-use -disorders	F19.96	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced persisting amnestic disorder
Substance-use -disorders	F19.97	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced persisting dementia
Substance-use -disorders	F19.980	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced anxiety disorder
Substance-use -disorders	F19.981	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced sexual dysfunction
Substance-use -disorders	F19.982	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced sleep disorder

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F19.988	ICD-10	Other psychoactive substance use, unspecified with other psychoactive substance-induced disorder
Substance-use -disorders	F19.99	ICD-10	Other psychoactive substance use, unspecified with unspecified psychoactive substance-induced disorder

For peer review only

Online supplementary material

eTable 7: U.S. State of Residence

State of residence	Full name	Number of patients	Percentage
CO	Colorado	125,267	23.0%
NY	New York	103,016	18.9%
MO	Missouri	98,450	18.1%
NC	North Carolina	75,626	13.9%
IN	Indiana	52,699	9.7%
VA	Virginia	23,848	4.4%
KY	Kentucky	18,102	3.3%
VT	Vermont	8,473	1.6%
DE	Delaware	7,313	1.3%
AK	Alaska	4,776	0.9%
MT	Montana	4,696	0.9%
AL	Alabama	3,459	0.6%
NJ	New Jersey	2,074	0.4%
IL	Illinois	1,994	0.4%
SC	South Carolina	1,193	0.2%
CT	Connecticut	1,125	0.2%
KS	Kansas	1,007	0.2%
FL	Florida	932	0.2%
NH	New Hampshire	695	0.1%
CA	California	679	0.1%
GA	Georgia	536	0.1%
PA	Pennsylvania	506	0.1%
TX	Texas	473	0.1%
MD	Maryland	451	0.1%
TN	Tennessee	448	0.1%
OH	Ohio	309	0.1%
WV	West Virginia	297	0.1%
MI	Michigan	272	0.1%
MA	Massachusetts	258	0.0%
WY	Wyoming	213	0.0%
MN	Minnesota	199	0.0%
AZ	Arizona	170	0.0%
WA	Washington	147	0.0%
NM	New Mexico	135	0.0%
WI	Wisconsin	111	0.0%
NE	Nebraska	98	0.0%
OK	Oklahoma	90	0.0%
LA	Louisiana	88	0.0%
IA	Iowa	82	0.0%
AR	Arkansas	78	0.0%

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

OR	Oregon	77	0.0%
SD	South Dakota	73	0.0%
UT	Utah	71	0.0%
DC	District of Columbia	67	0.0%
ND	North Dakota	63	0.0%
NV	Nevada	60	0.0%
MS	Mississippi	53	0.0%
RI	Rhode Island	42	0.0%
ME	Maine	42	0.0%
PR	Puerto Rico	30	0.0%
ID	Idaho	26	0.0%
HI	Hawaii	23	0.0%
AE	Armed Forces Europe	<20	0.0%
AP	Armed Forces Pacific	<20	0.0%

For peer review only

Online supplementary material

eTable 8: Mean CGI-S score by diagnosis

Diagnosis	Number of patients	Mean CGI-S	95% Confidence Interval	Standard deviation
Dementia & Alzheimer's disease	5,029	4.33	4.30 to 4.37	1.36
Substance-use disorders	152,790	3.61	3.60 to 3.62	1.80
Schizophrenia	14,592	4.44	4.41 to 4.46	1.44
Schizoaffective disorder	15,044	4.48	4.45 to 4.50	1.42
Bipolar disorder	69,607	4.28	4.27 to 4.29	1.45
Major depressive disorder (MDD)	129,120	4.21	4.20 to 4.22	1.42
Anxiety disorders	103,923	4.02	4.01 to 4.03	1.53
Attention-deficit hyperactivity disorder (ADHD)	53,744	3.99	3.97 to 4.00	1.42
Note: Confidence intervals were estimated using the formula mean +/- error margin where: Error margin = std/sqrt(N) * Z, Z = 1.96 for 95% confidence interval				

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

eTable 9: Mean maximum and minimum CGI-S score by diagnosis

Diagnosis	Number of patients with at least 2 CGI-S data points	Max CGI-S Mean	Min CGI-S Mean	Max CGI-S Std dev	Min CGI-S Std dev	Max CGI-S Confidence Interval	Min CGI-S Confidence Interval
Dementia & Alzheimer's disease	2,980	5.10	3.32	1.06	1.56	5.06 to5.14	3.26 to 3.37
Substance-use disorders	107,939	4.85	2.55	1.32	1.62	4.84 to 4.85	2.54 to 2.56
Schizophrenia	12,465	5.41	2.90	1.01	1.55	5.40 to 5.43	2.87 to 2.93
Schizoaffective disorder	13,590	5.53	2.88	0.95	1.55	5.51 to 5.55	2.86 to 2.91
Bipolar disorder	59,222	5.22	3.02	0.98	1.54	5.21 to 5.23	3.00 to 3.03
Major depressive disorder (MDD)	103,933	5.02	3.03	0.98	1.53	5.01 to 5.03	3.03 to 3.04
Anxiety disorders	89,710	5.04	2.80	0.97	1.58	5.03 to 5.04	2.79 to 2.81
Attention-deficit hyperactivity disorder (ADHD)	47,591	4.92	2.75	1.00	1.50	4.91 to 4.93	2.73 to 2.76

Note: Confidence intervals were estimated using the formula mean +/- error margin where:
Error margin = std/sqrt(N) * Z, Z = 1.96 for 95% confidence interval

For peer review only

Online supplementary material

eTable 10: Number of patients with at least one hospital visit by diagnosis

Diagnosis	Number of patients	Number of hospitalized patients	Percentage
Dementia & Alzheimer's disease	5,029	2,625	52.2%
Substance-use disorders	152,790	54,267	35.5%
Schizophrenia	14,592	8,123	55.7%
Schizoaffective disorder	15,044	8,528	56.7%
Bipolar disorder	69,607	25,322	36.4%
Major depressive disorder (MDD)	129,120	40,726	31.5%
Anxiety disorders	103,923	22,455	21.6%
Attention-deficit hyperactivity disorder (ADHD)	53,744	8,587	16.0%
Note: Hospitalization includes both Inpatient and Emergency services			

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

eTable 11: SAV features included in MSE NLP analysis (eTable 12)

Category	Description	SAV number
Delusions	experiencing delusions/abnormal thoughts (grandeur)	sav_1_aopt_1
Delusions	experiencing delusions/abnormal thoughts (persecution)	sav_2_aopt_2
Delusions	experiencing delusions/abnormal thoughts (religious)	sav_3_aopt_3
Delusions	experiencing hallucinations (auditory)	sav_4_aopt_4
Delusions	experiencing hallucinations (visual)	sav_5_aopt_5
Delusions	experiencing hallucinations (tactile)	sav_6_aopt_6
Delusions	History of delusions/abnormal thoughts	sav_7_aopt_7
Delusions	History of hallucinations	sav_8_aopt_8
Delusions	experiencing delusions (not specified)	sav_10_aopt_10
Delusions	experiencing hallucinations (not specified)	sav_11_aopt_11
Delusions	experiencing delusions/abnormal thoughts (paranoia)	sav_13_aopt_13
Delusions	experiencing hallucinations (olfactory)	sav_14_aopt_14
Delusions	responding to internal stimuli	sav_15_aopt_15
Delusions	experiencing delusions/abnormal thoughts (sexual)	sav_16_aopt_16
Mood	"anxious, tense"	sav_158_mood_2
Mood	"depressed, sad, despondent"	sav_159_mood_3
Mood	"irritable, angry"	sav_160_mood_4
Mood	declined	sav_162_mood_6
Mood	"elevated, manic"	sav_163_mood_7
Mood	labile	sav_164_mood_8
Mood	misc. issues	sav_165_mood_9
Mood	suicidal	sav_166_mood_10
Cognition	general issues	sav_65_attn_concentration_2
Cognition	declined	sav_67_attn_concentration_4
Cognition	issues due to MH	sav_69_attn_concentration_6
Cognition	issues due to DD	sav_70_attn_concentration_7
Cognition	varying	sav_71_attn_concentration_8
Cognition	issues with concentration	sav_81_cognition_2
Cognition	issues with fund of knowledge	sav_83_cognition_4
Cognition	issues with attention	sav_85_cognition_6
Cognition	general issues	sav_87_cognition_8
Cognition	some impairment	sav_90_exec_functioning_2
Cognition	much impairment	sav_91_exec_functioning_3
Cognition	declined since last session	sav_93_exec_functioning_5
Cognition	below age level	sav_95_exec_functioning_7
Cognition	impaired (due to DD)	sav_96_exec_functioning_8
Cognition	impaired (due to TBI)	sav_97_exec_functioning_9
Cognition	impaired (due to other MH issue)	sav_98_exec_functioning_10

For peer review only

Online supplementary material

eTable 12: Number of patients with at least one MSE feature by diagnosis

Diagnosis	MSE Category	Number of patients	Number of patients with at least one MSE feature	Percentage
Dementia & Alzheimer's disease	Delusions & hallucinations	5,029	2,147	42.7%
Dementia & Alzheimer's disease	Mood	5,029	2,518	50.1%
Dementia & Alzheimer's disease	Cognition	5,029	2,719	54.1%
Substance-use disorders	Delusions & hallucinations	152,790	37,455	24.5%
Substance-use disorders	Mood	152,790	89,616	58.7%
Substance-use disorders	Cognition	152,790	26,461	17.3%
Schizophrenia	Delusions & hallucinations	14,592	11,513	78.9%
Schizophrenia	Mood	14,592	10,753	73.7%
Schizophrenia	Cognition	14,592	6,418	44.0%
Schizoaffective disorder	Delusions & hallucinations	15,044	11,885	79.0%
Schizoaffective disorder	Mood	15,044	12,278	81.6%
Schizoaffective disorder	Cognition	15,044	6,785	45.1%
Bipolar disorder	Delusions & hallucinations	69,607	33,179	47.7%
Bipolar disorder	Mood	69,607	55,472	79.7%
Bipolar disorder	Cognition	69,607	23,373	33.6%
Major depressive disorder (MDD)	Delusions & hallucinations	129,120	42,566	33.0%
Major depressive disorder (MDD)	Mood	129,120	100,182	77.6%
Major depressive disorder (MDD)	Cognition	129,120	28,986	22.4%
Anxiety disorders	Delusions & hallucinations	103,923	32,059	30.8%
Anxiety disorders	Mood	103,923	76,320	73.4%
Anxiety disorders	Cognition	103,923	25,625	24.7%
Attention-deficit hyperactivity disorder (ADHD)	Delusions & hallucinations	53,744	11,933	22.2%
Attention-deficit hyperactivity disorder (ADHD)	Mood	53,744	31,112	57.9%
Attention-deficit hyperactivity disorder (ADHD)	Cognition	53,744	18,432	34.3%

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

eFigure 1: MindLinc EHR data de-identification and harmonization pipeline

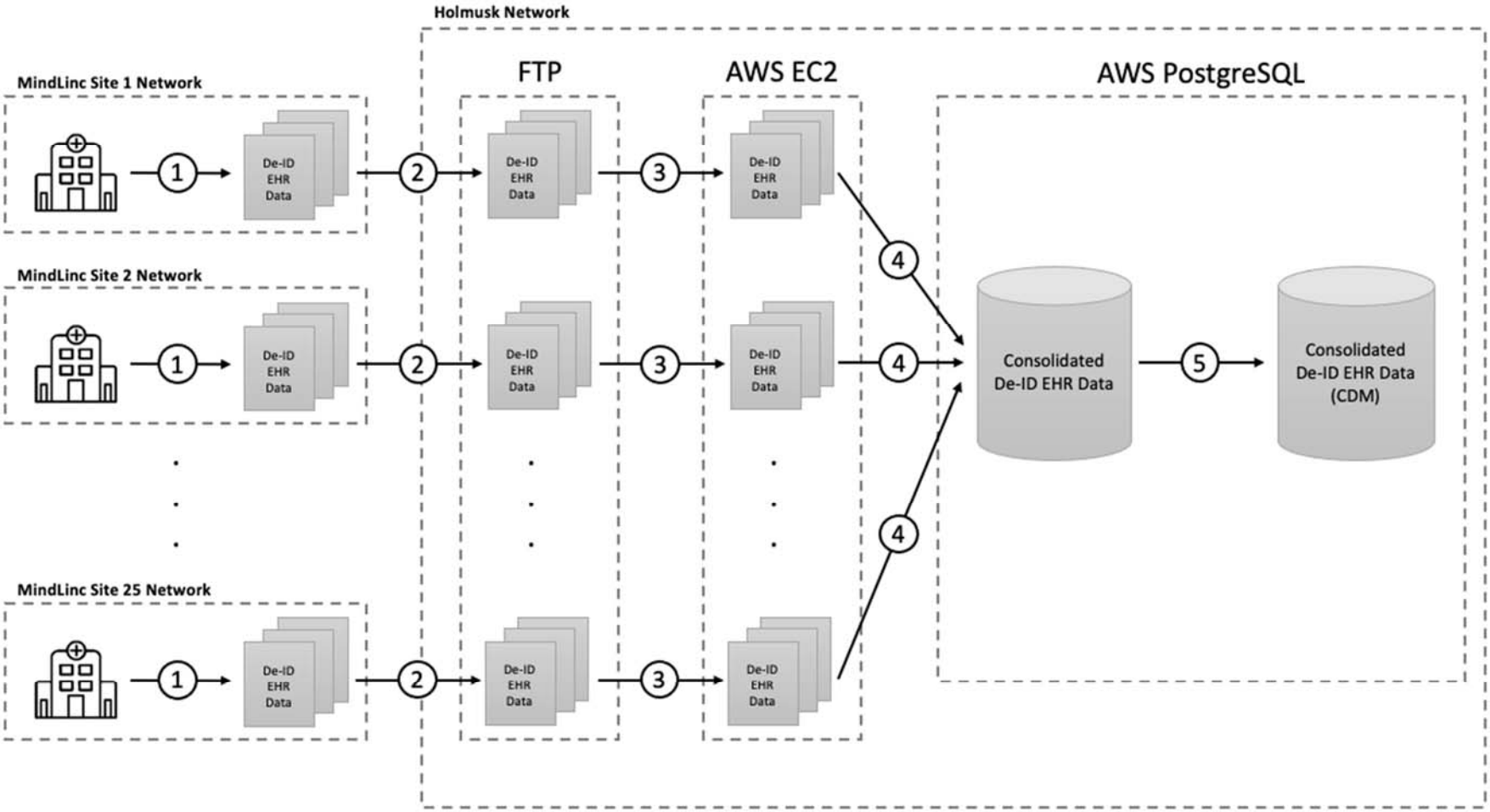


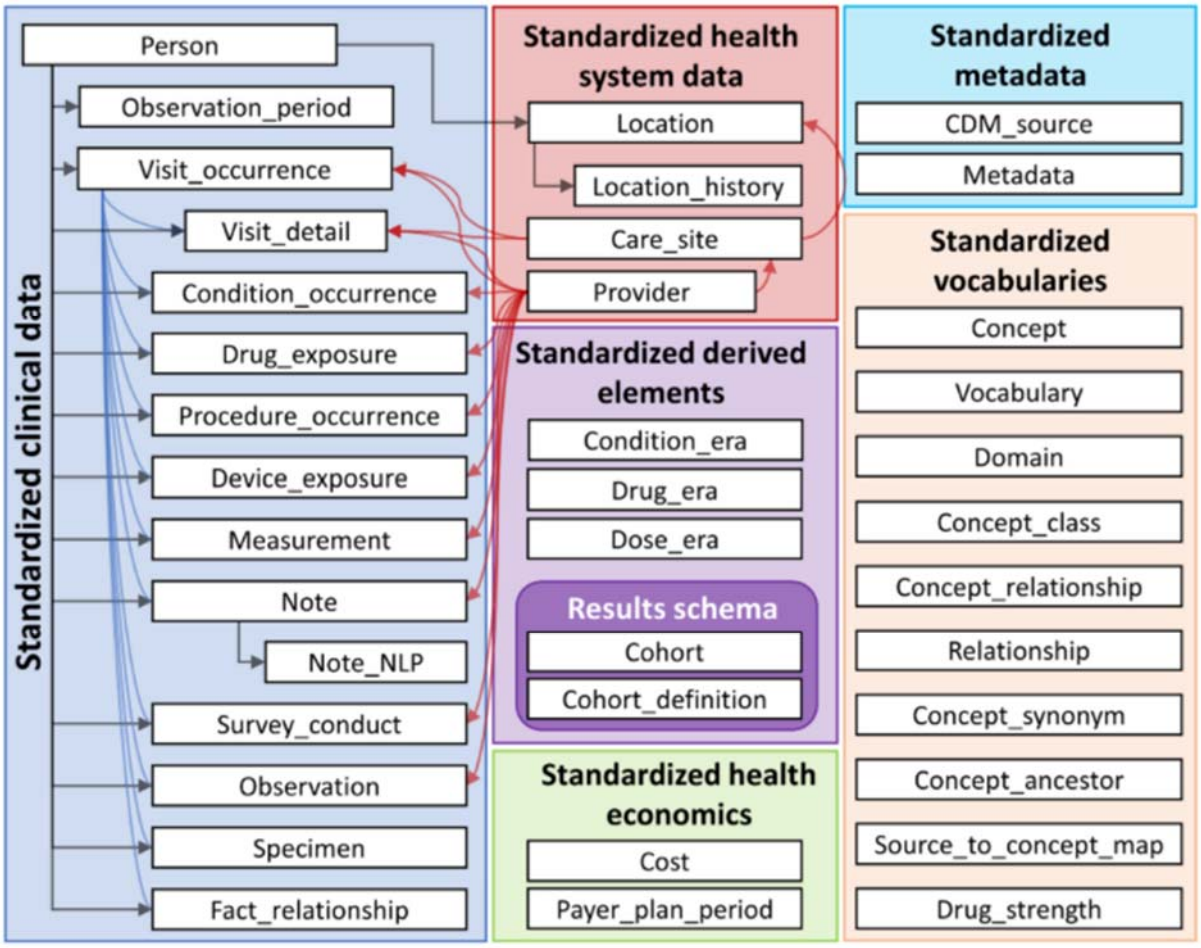
Figure: MindLinc Data Flow Diagram

Online supplementary material

eFigure 2: OMOP CDM schematic

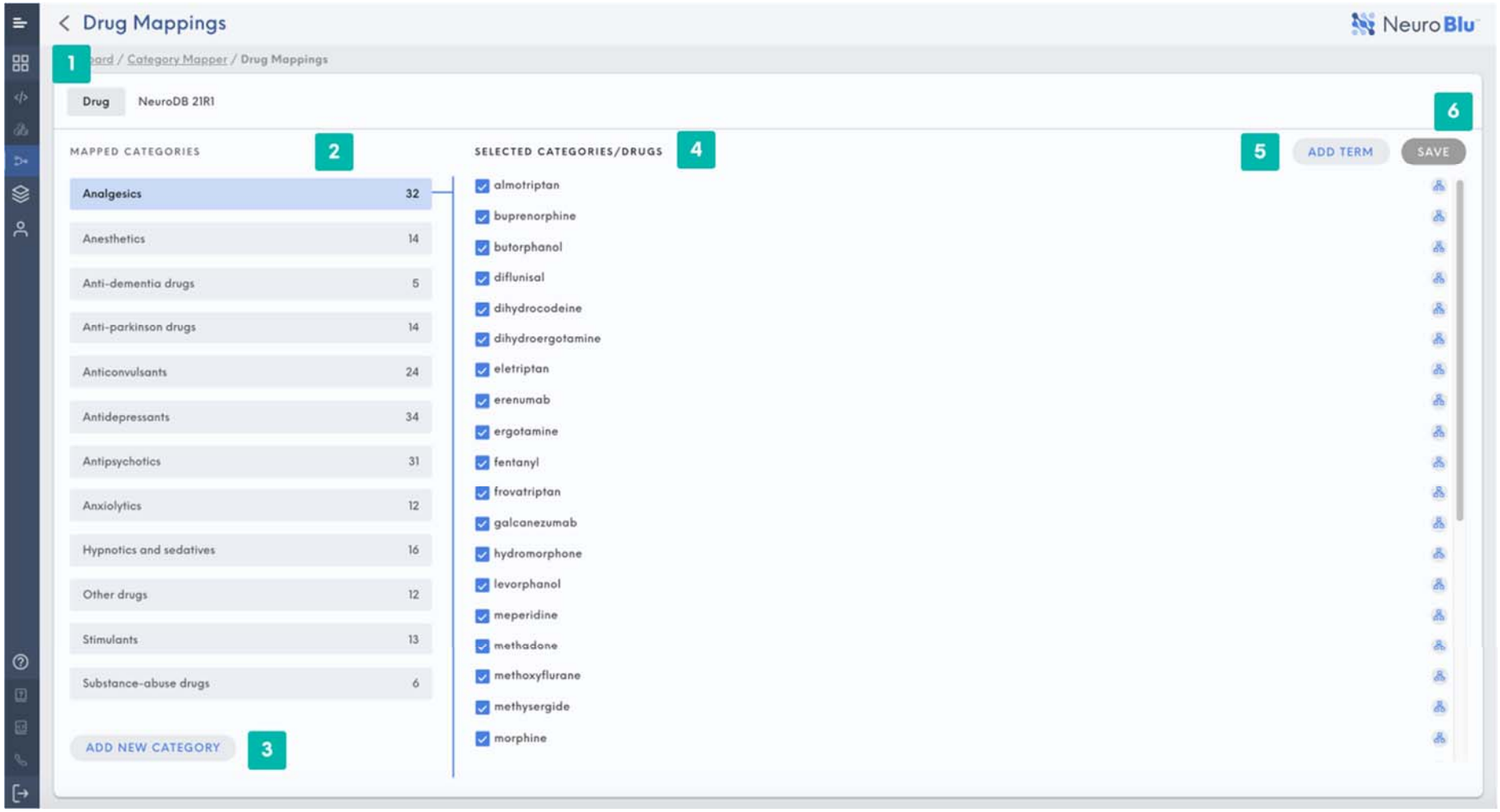
Source: https://ohdsi.github.io/TheBookOfOhdsi/CommonDataModel.html

Attribution: Creative Commons Zero v1.0 Universal license (No copyright)



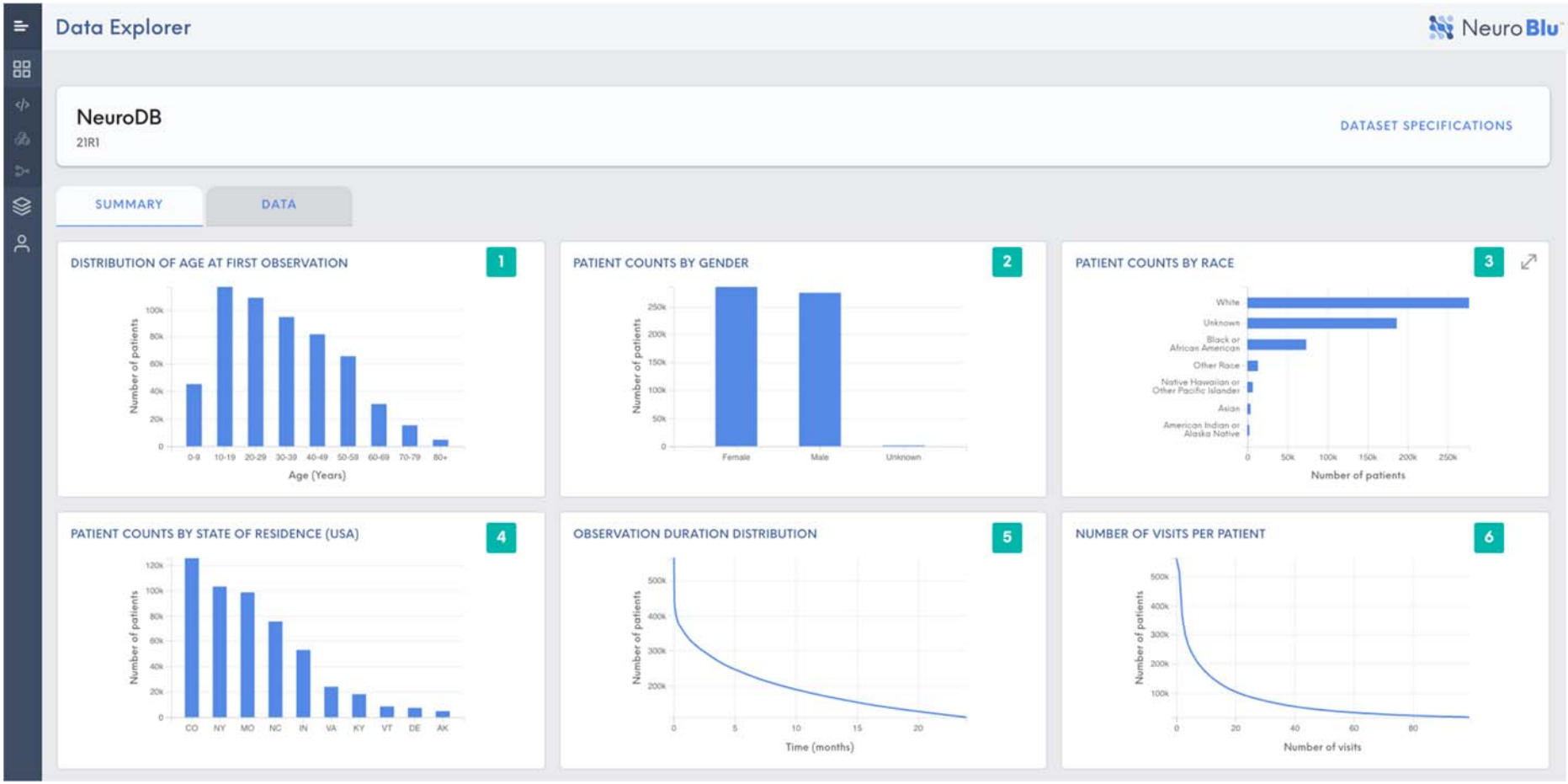
Online supplementary material

eFigure 3: NeuroBlu Category Mapper illustrating the classification of individual drugs into different types of medication

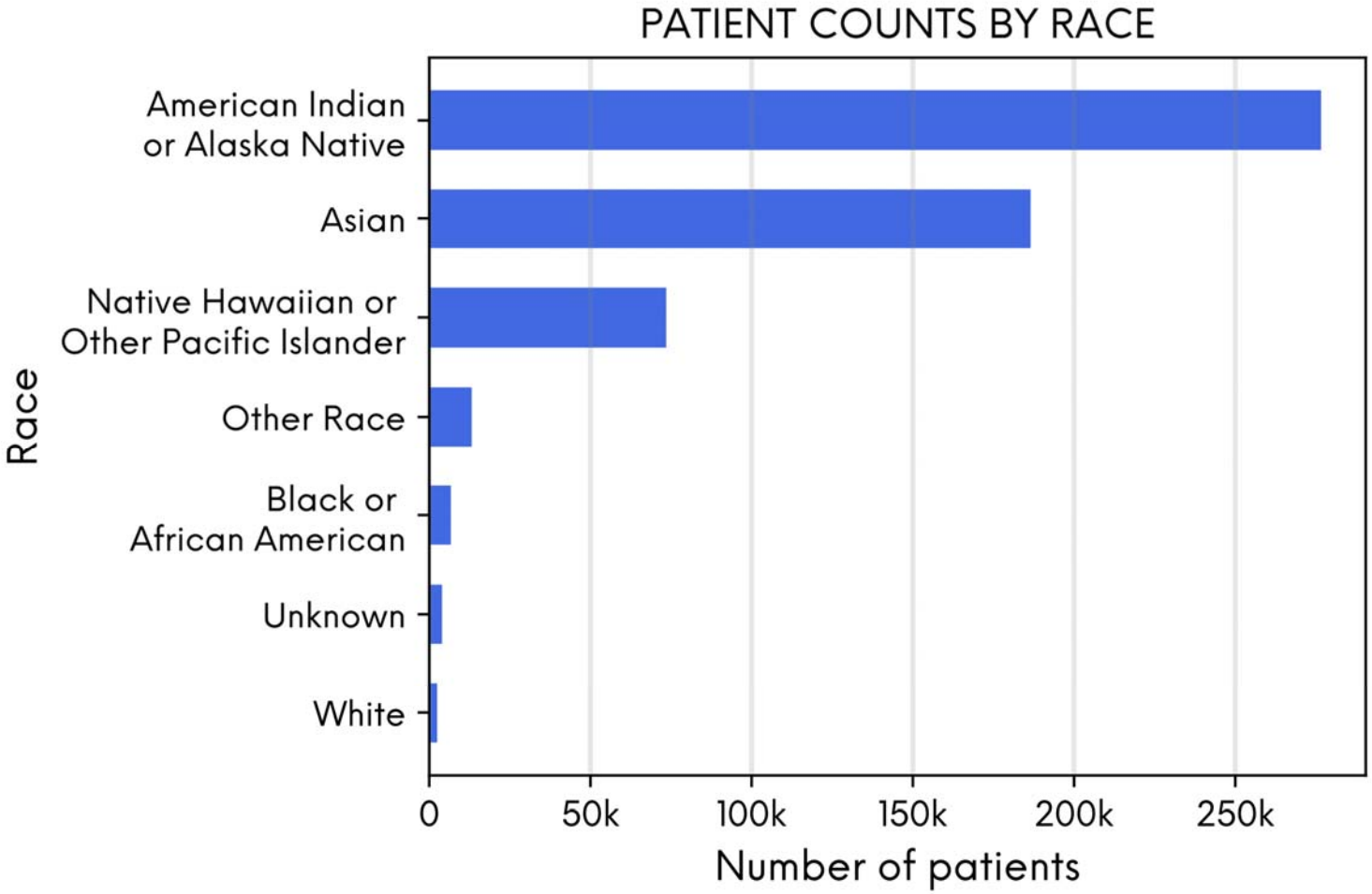


Online supplementary material

eFigure 4: NeuroBlu Data Explorer which provides a summary of the dataset and allows the top 50 rows of each table to be viewed

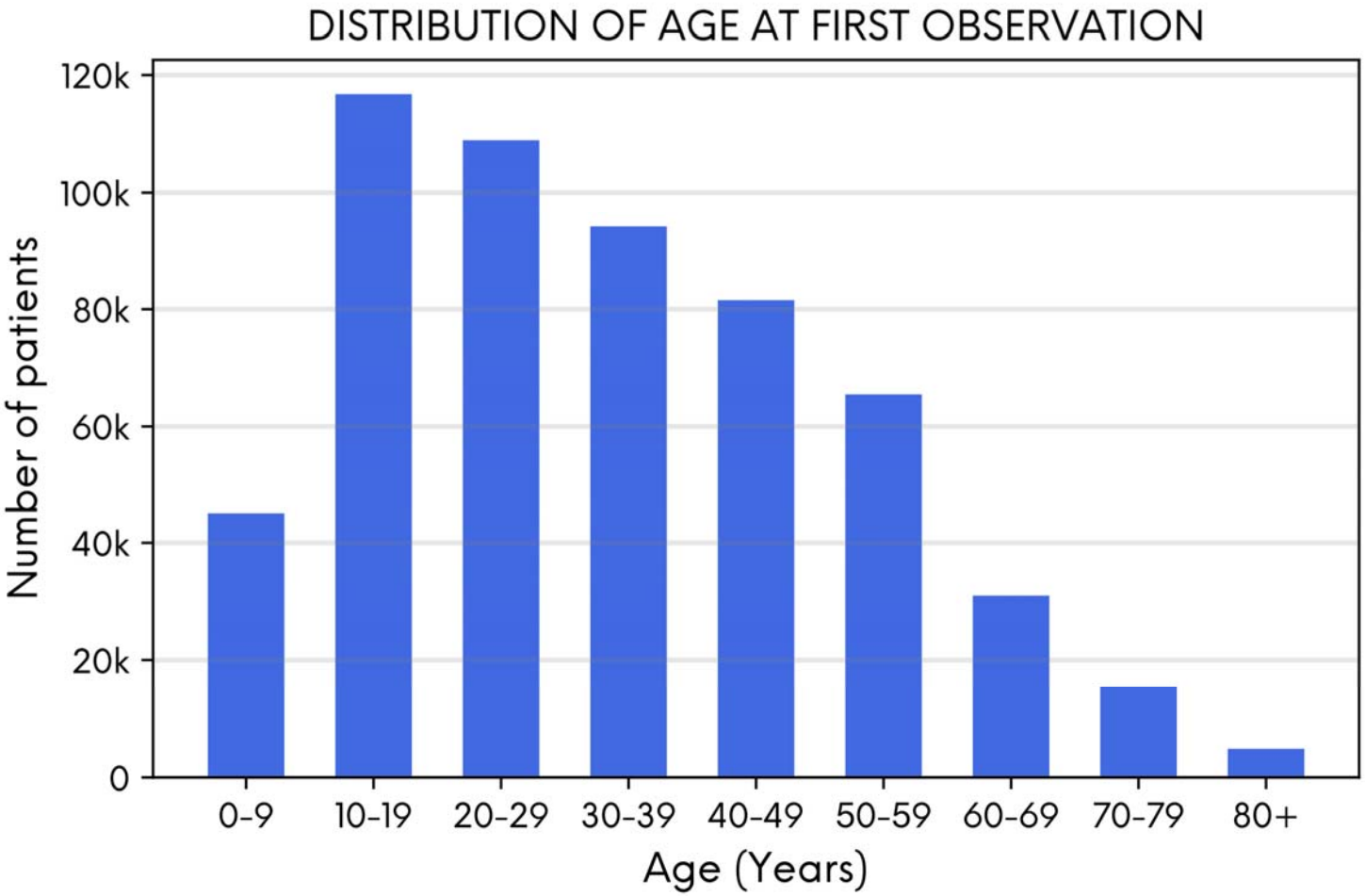


eFigure 5: NeuroBlu dataset race distribution

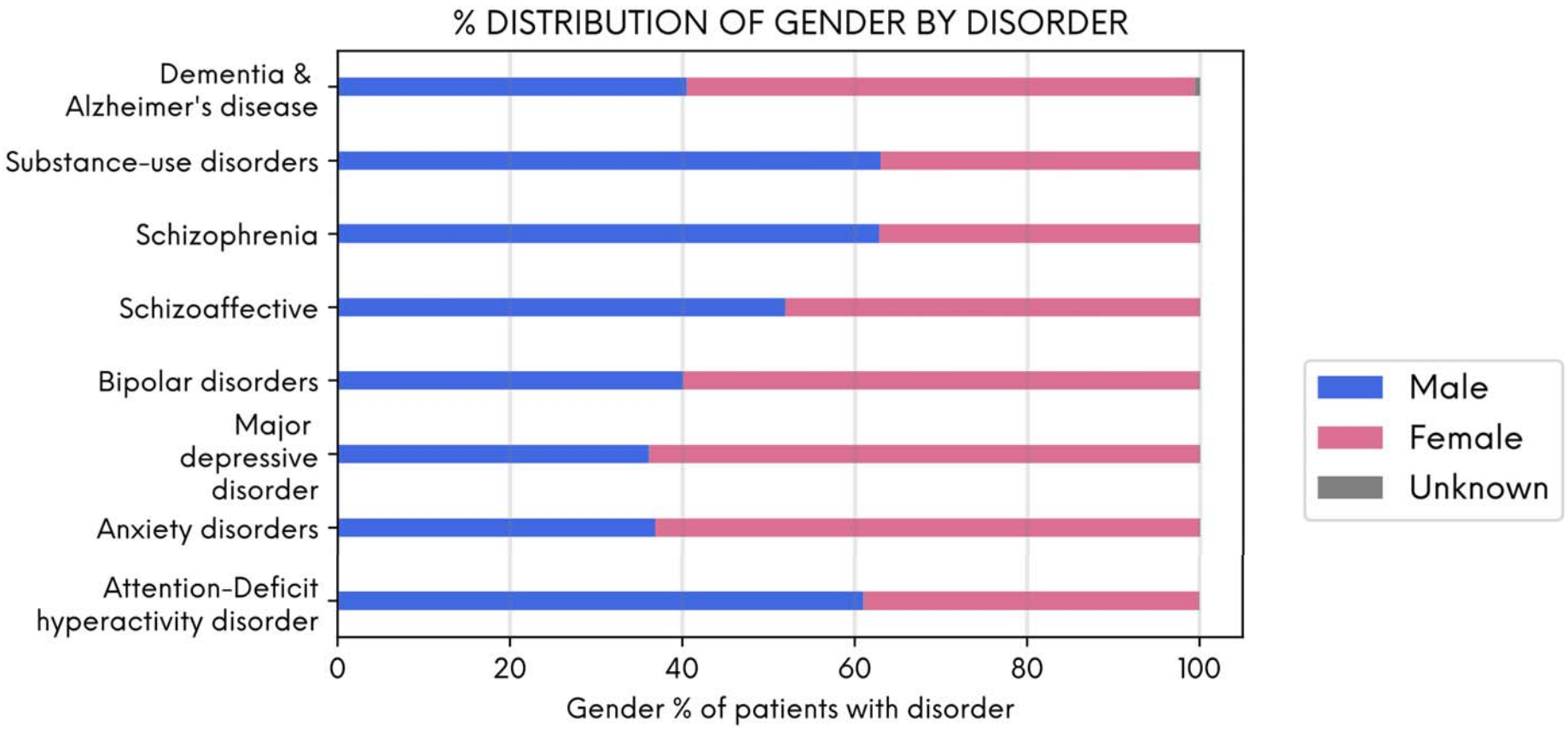


Online supplementary material

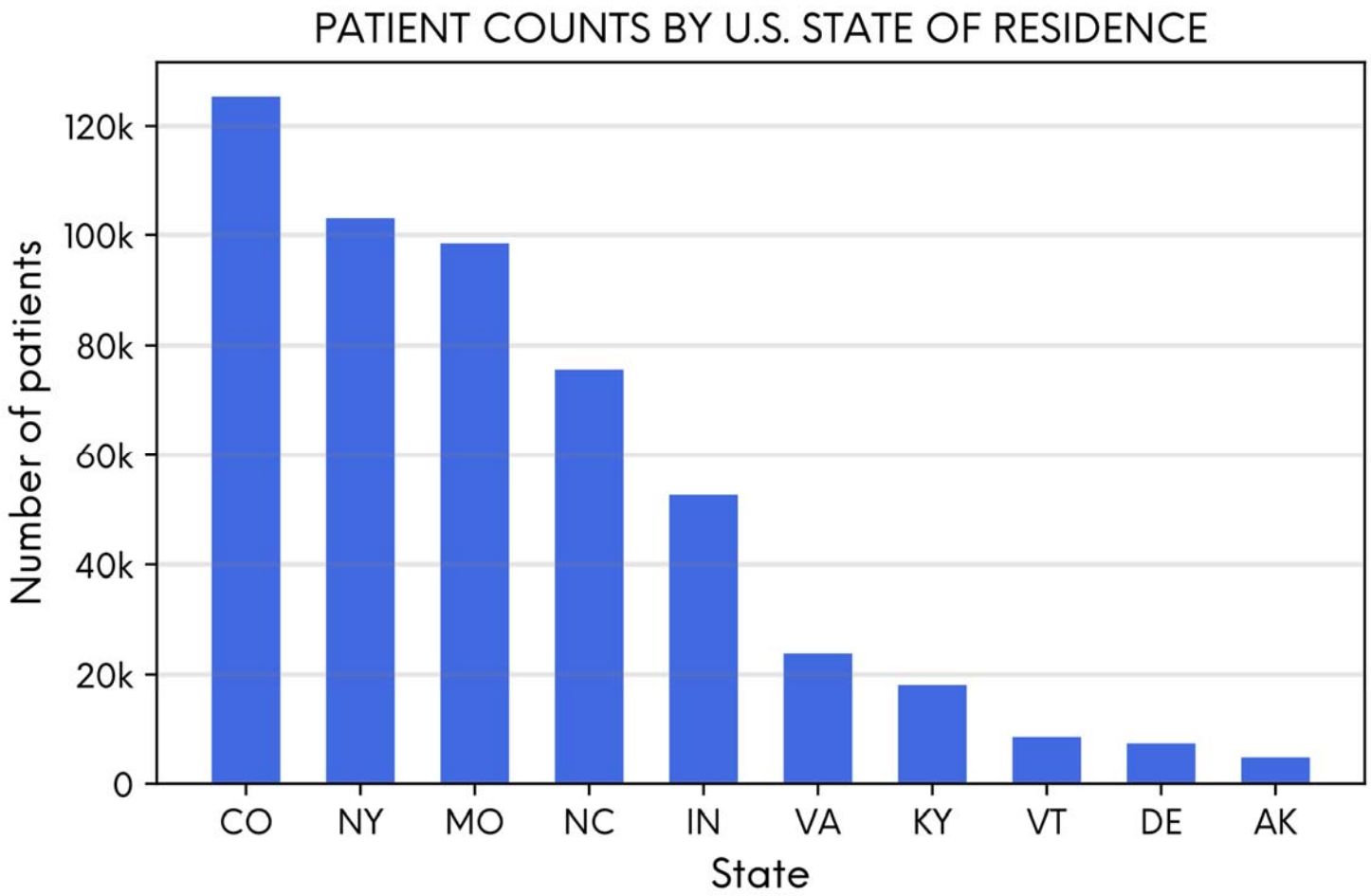
eFigure 6: NeuroBlu dataset age distribution



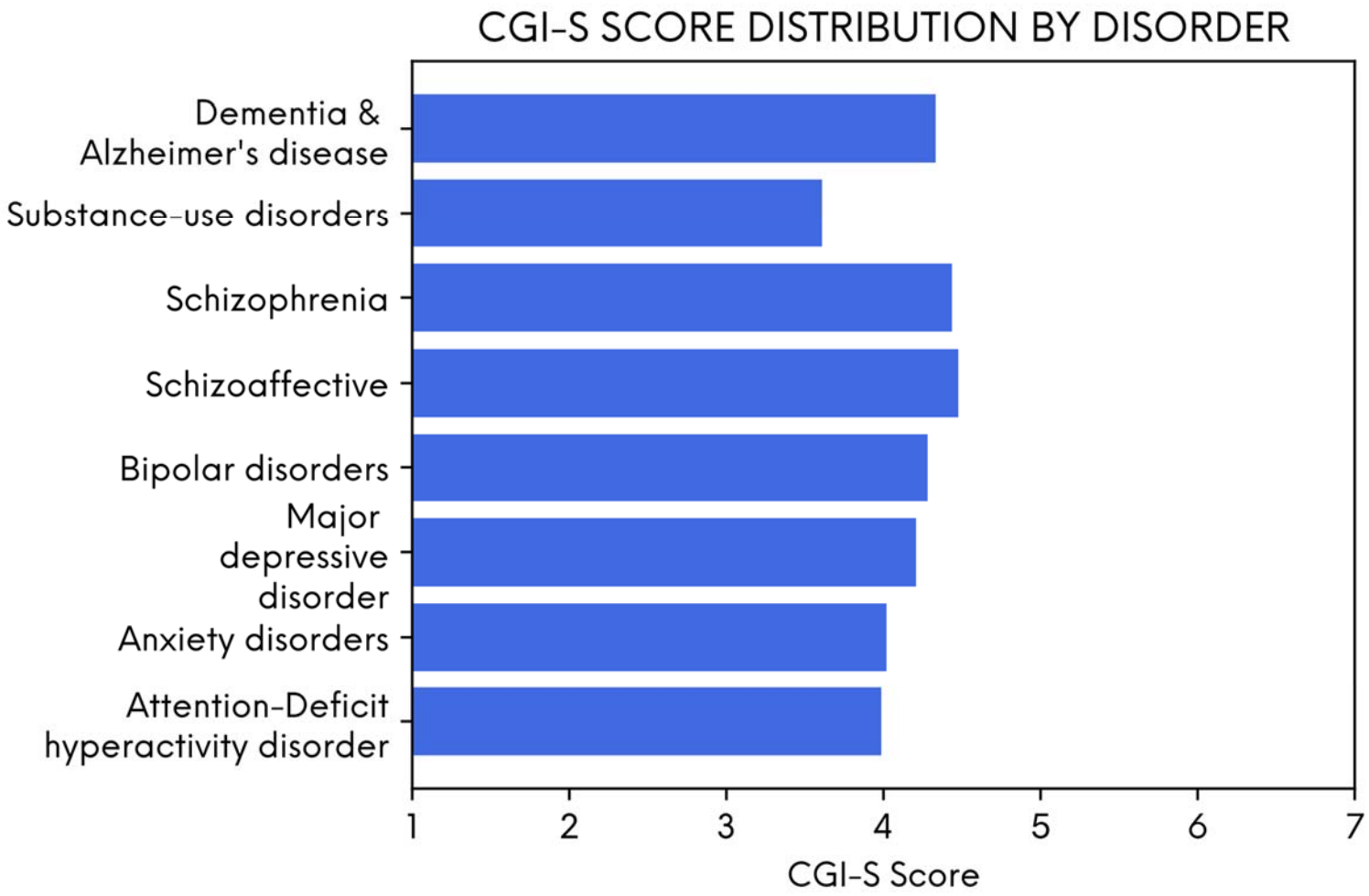
eFigure 7: Gender distribution by diagnosis



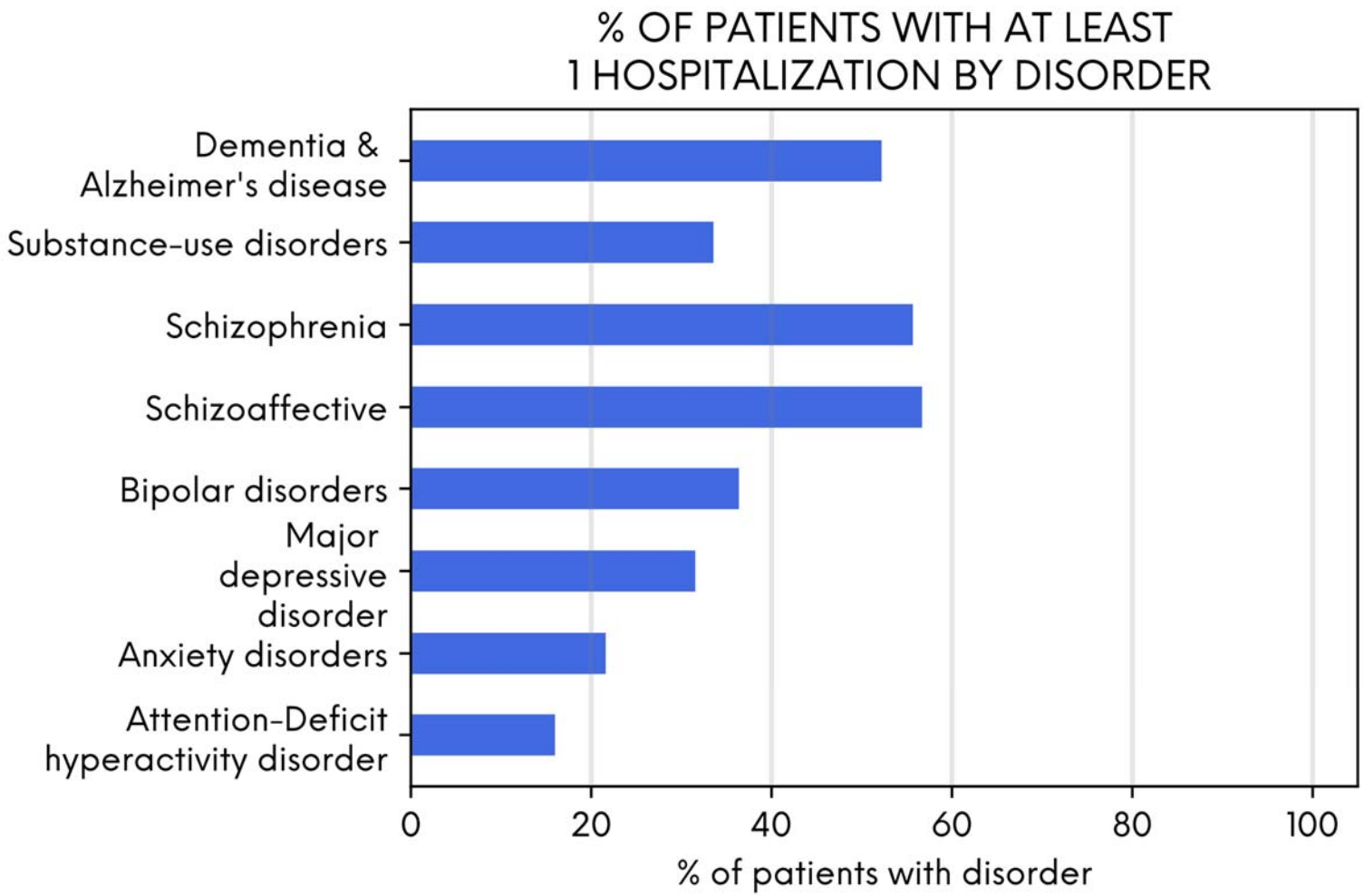
eFigure 8: U.S. State of Residence



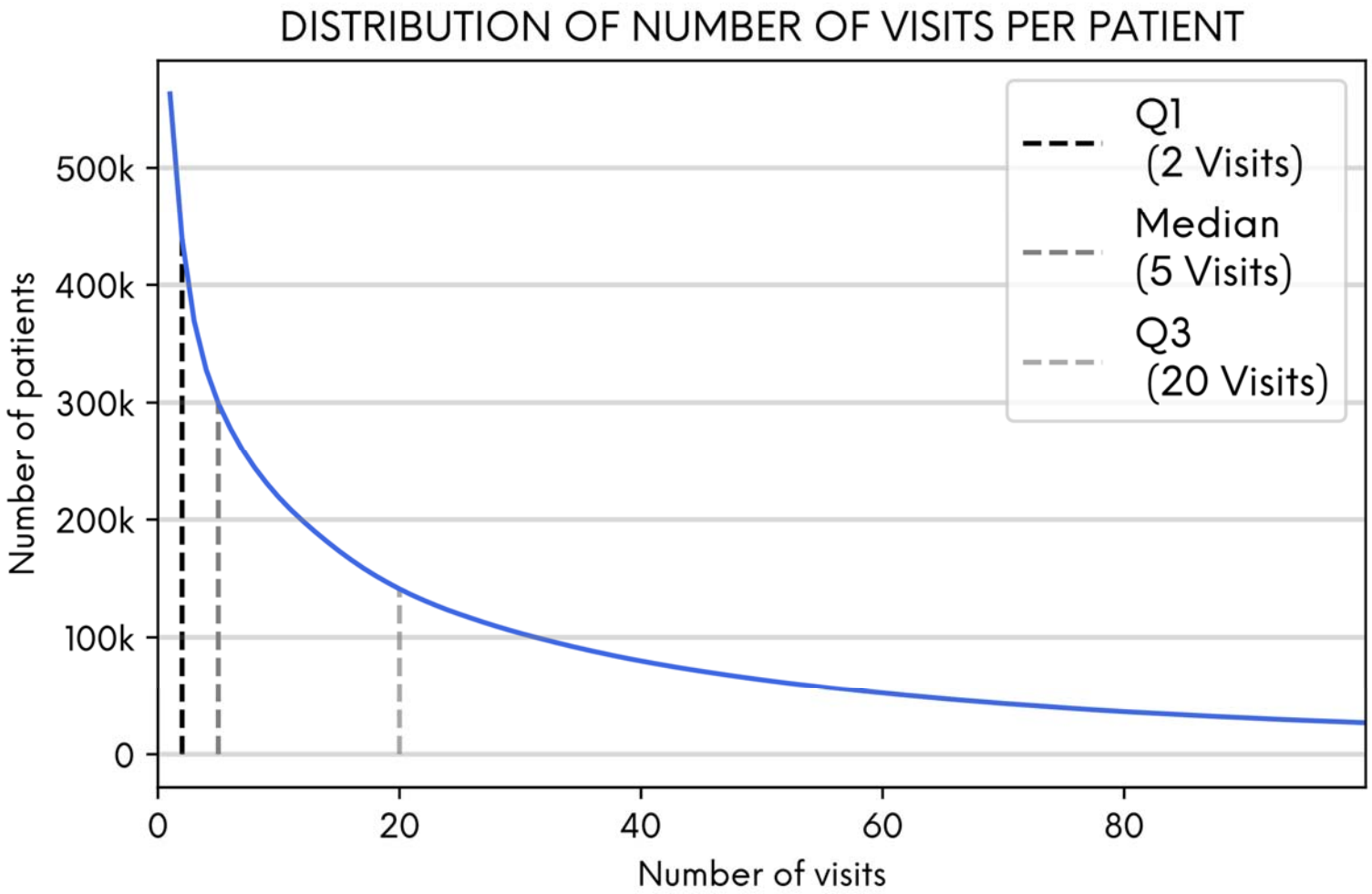
eFigure 9: Mean CGI-S score by diagnosis



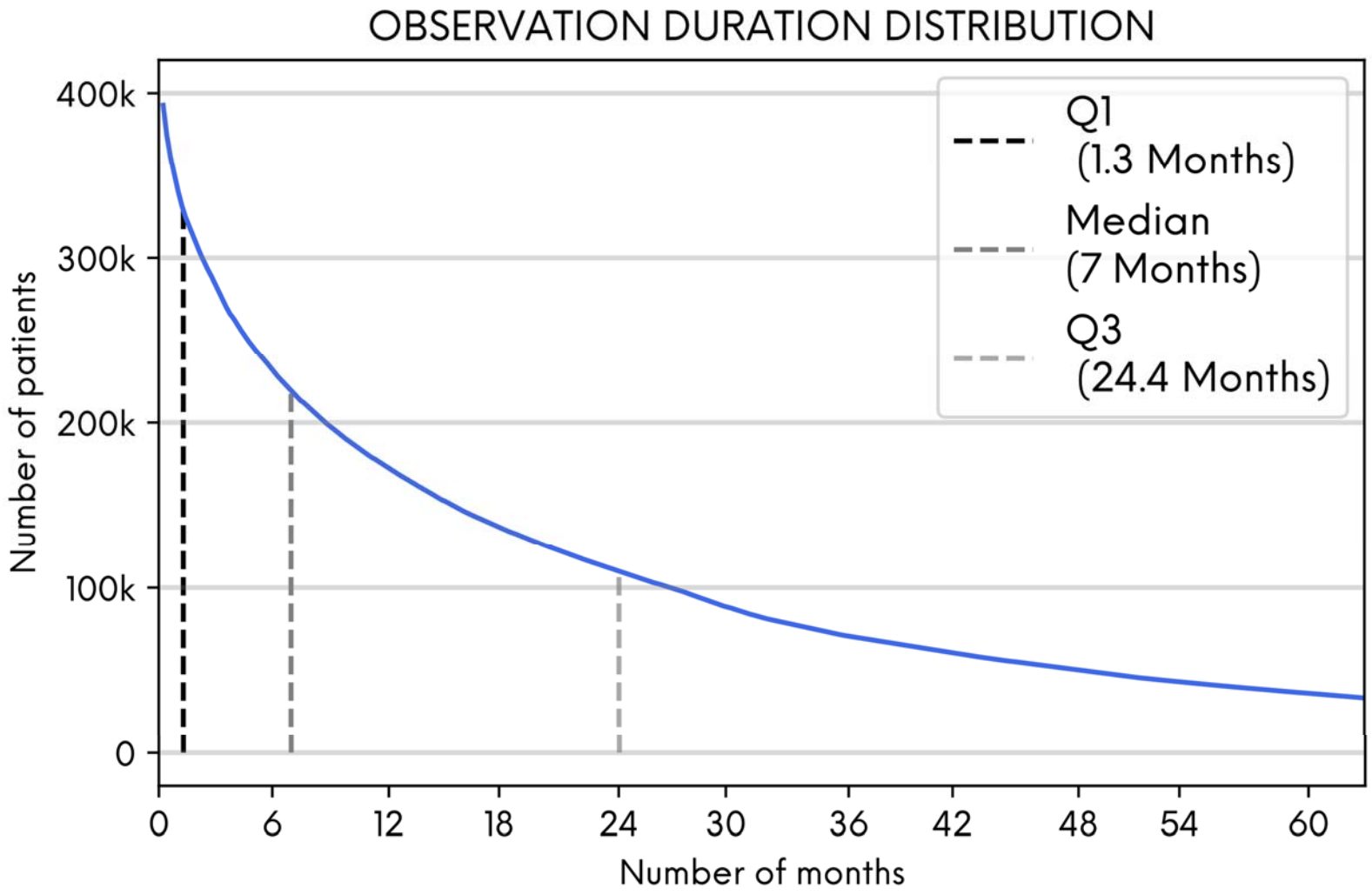
eFigure 10: Percentage of patients with at least one hospital visit by diagnosis



eFigure 11: Distribution of number of visits per patient



eFigure 12: Distribution of duration of follow-up



BMJ Open

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare analytics with real-world data

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-057227.R2
Article Type:	Cohort profile
Date Submitted by the Author:	10-Mar-2022
Complete List of Authors:	Patel, Rashmi; Holmusk USA Inc; King's College London Institute of Psychiatry Psychology and Neuroscience, Department of Psychosis Studies Wee, Soon Nan; Holmusk USA Inc Ramaswamy, Rajagopalan; Holmusk USA Inc Thadani, Simran; Holmusk USA Inc Tandi, Jesisca; Holmusk USA Inc Garg, Ruchir; Holmusk USA Inc Calvanese, Nathan; Holmusk USA Inc Valko, Matthew; Holmusk USA Inc Rush, A.; Curbstone Consultant LLC Rentería, Miguel; Holmusk USA Inc Sarkar, Joydeep; Holmusk USA Inc Kollins, Scott; Holmusk USA Inc; Duke University School of Medicine
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Health informatics
Keywords:	PSYCHIATRY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, EPIDEMIOLOGY

SCHOLARONE™
Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

For peer review only

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment 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] and Scott H. Kollins[1,4]

1. Holmusk, 54 Thompson Street, New York, NY 10012, USA
2. King's College London (Institute of Psychiatry, Psychology and Neuroscience), London, UK
3. Curbstone Consultant LLC, Santa Fe, NM, USA
4. Duke University School of Medicine (Department of Psychiatry and Behavioral Sciences), Durham, NC, USA

Correspondence to:

Rashmi Patel

Holmusk, 54 Thompson Street, New York, NY 10012, USA

E-mail: rashmi.patel@holmusk.com

Manuscript word count (excluding references): 5,656

Abstract word count: 298

For peer review only

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare real-world data analytics

Abstract

Purpose:

NeuroBlu is a real-world data (RWD) repository that contains de-identified electronic health record (EHR) data from U.S. 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, ICD-9/ICD-10 diagnosis, prescribed medications, family history of mental disorders, Clinical Global Impression (CGI-S/CGI-I), and Global Assessment of Functioning (GAF). To further enhance the dataset, 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 dataset supports the generation of real-world evidence (RWE) in mental health.

Participants:

As of 31st July 2021, 562,940 individuals (48.9% male) were present in the dataset with a mean age of 33.4 years (standard deviation: 18.4 years). The most frequently recorded diagnoses were substance use disorders (152,790 patients), major depressive disorder (129,120 patients) and anxiety disorders (103,923 patients). The median duration of follow-up was 7 months (interquartile range: 1.3 to 24.4 months).

Findings to date:

The dataset has supported epidemiological studies demonstrating increased risk of psychiatric hospitalization and reduced antidepressant treatment effectiveness among people with comorbid substance use disorders. It has also been used to develop data visualization 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 dataset will be further analyzed 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 health care services.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare real-world data analytics

Article Summary

Strengths and limitations

- The NeuroBlu dataset benefits from a large sample size of de-identified electronic health record (EHR) data from over 560,000 people who have received mental healthcare over a period of 21 years.
- The dataset is built on a robust de-identification pipeline and encryption framework that enables a wide range of users to safely analyze data through a trusted research environment using a graphical user interface or advanced analytic software (R and Python).
- Structured data on clinical severity (CGI-S) are recorded for over 80% of patients in the dataset. Natural language processing (NLP) enables access to rich clinical data from unstructured free text data as part of the Mental State Examination (MSE) and social factors that are not typically available in purely structured clinical datasets such as claims or RCT data.
- As the dataset 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.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 U.S. 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 expectancy[3] and the lifetime costs associated with SMI are estimated to be $1.85 million 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 datasets, known as real world data (RWD), has improved the understanding of a range of medical disorders at a population level[7] and contributed to the development[8] and evaluation[9] of novel therapeutics. RWD may be analyzed to generate real-world evidence (RWE) to quantify the safety and effectiveness of therapeutic interventions. The generation of RWE from the analysis of largescale datasets that are representative of healthcare service delivery in routine settings provides a complementary source of data – as compared to more controlled clinical trials – that 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 characterize health disorders and to develop and deliver more effective treatments.[12] EHR datasets have previously supported RWE generation in critical care[13] and infrastructure based on common data models[14] has been developed to enable researchers to securely analyze 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-standardized 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 analyzed. Relatedly, there is little standardization 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 de-identified mental health EHR datasets have been curated to support largescale population research,[19–21] provide clinical phenotype data for genomics research,[22] 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

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

unstructured free text in EHRs and enable statistical analysis of detailed clinical data such as presenting symptoms,[27] environmental factors,[28] 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 datasets are only accessible by clinical researchers based in specialized academic health science centers.[31] Enabling wider access, while maintaining robust data privacy and security standards, could substantially increase the impact of de-identified EHR datasets 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 de-identified EHR data from U.S. mental healthcare providers, including state-of-the-art NLP software to characterize mental state information from semi-structured 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.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Cohort Description

Participants and setting

The NeuroBlu platform enables users to analyze de-identified data from over 560,000 patients receiving care from 25 U.S. mental healthcare providers that operate the MindLinc EHR system. Supplementary eTable 1 provides details of the locations, types of mental healthcare service and numbers of patients for each provider. In summary, the platform includes de-identified EHR data from individuals receiving mental healthcare from outpatient, inpatient, telemedicine, and residential care facilities in 12 U.S. 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 semi-structured free text fields to document the mental state examination (MSE) and treatment plan.[33] A subset of de-identified 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 de-identified at source before being transformed and normalized in a cloud-based U.S. Amazon Web Services (AWS) data warehouse and mapped to a common data model to create a harmonized dataset including data from all participating healthcare providers (eFigure 1). NeuroBlu users analyze the harmonized dataset 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, de-identified EHR data may be analyzed by anyone with access to the trusted research environment without requiring any movement or disclosure of the underlying data.

De-identification procedure

MindLinc EHR data are de-identified at source (i.e. within the computing infrastructure of each participating healthcare provider) using the Safe Harbor method (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 dataset. 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, URLs, social security numbers, 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 randomizing identifiers, recording age in years, and collecting the first three digits of zip codes with a population greater than 20,000).

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Transfer of de-identified EHR data from healthcare provider to secure AWS cloud

The de-identified EHR data are compressed and transferred from the source healthcare provider through a virtual private network (VPN) to a secure file transfer protocol (FTP) site (eFigure 1, step 2). The data are then transferred to a secure Amazon Elastic Compute Cloud (AWS EC2) instance (eFigure 1, step 3) where they are decompressed and assembled for loading into a PostgreSQL database (eFigure 1, step 4).

Given the relational nature of the MindLinc databases, the files produced by the EHR data de-identification 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 de-identification process is irreversible, preventing re-identification 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 (eFigure 1, step 4) and is further described in section 1 of the Supplementary 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 (e.g. appearance and behavior) or may be elicited through direct questioning or the presentation of cognitive tasks. The MSE provides a rich source of clinical phenotype data which characterizes the nature of presenting symptoms which could be associated with varying clinical outcomes and response to treatment.

The MindLinc EHR includes a semi-structured "status assessment" field in which clinicians can document features associated with a patient's MSE. The status assessment field allows clinicians to choose pre-defined features from a list of options and/or to document findings as unstructured free text. However, pre-defined 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 pre-defined features in MindLinc is 0% whereas the median percentage with unstructured free text is 67%.

NLP is the sub-discipline 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: (i) Data assembly: identifying a collection of relevant documents (the corpus), (ii) 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, (iii) Preprocessing: preparation of the corpus for NLP model development including stop word removal, stemming, lemmatization and parts of speech tagging, (iv) Featurization: classifying text within the corpus into different features (e.g. parts of speech, word vectors or embeddings, sentiment or temporal features) and (v) Analysis: development of NLP algorithm using a rules-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 maximize accuracy as measured through precision (positive predictive value), recall (sensitivity) and F1 measure (harmonic mean of precision and recall).[37] If a sufficiently accurate models are developed, they can be applied to the entire corpus to generate structured data on clinically

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

meaningful features of interest[27] 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 de-identified EHR dataset to create a table of NLP-derived MSE measures which may be joined and analyzed with other structured clinical and sociodemographic data in the NeuroBlu dataset. 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 below in the section on *Unstructured data*.

Data variables

Full details of available data variables and their structure within the NeuroBlu MindLinc dataset are provided in section 2 of the Supplementary 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 dataset includes structured sociodemographic data on gender, year of birth, race, ethnicity, U.S. 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 dataset that include data on contacts with mental health services, emergency room visits, hospital admissions, 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 dataset. Structured questionnaire data are available for the following numbers of patients within the dataset:

(i) CGI-S: 471,256 patients (83.7%)
(ii) CGI-I: 361,819 patients (64.3%)
(iii) GAF: 310,895 patients (55.3%)
(iv) MADRS: 1,126 patients (0.2%)

The dataset 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 Supplementary Material.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 Supplementary 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 & MSE data, only the data derived using NLP for 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; Judgment; Language; Level of consciousness; Memory; Mood; Orientation; Psychomotor; Reasoning; Sensorium; Sleep; Speech; Suicidal; 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 (e.g. 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 analyzed 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 to 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 target 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: (i) Cohort Builder, (ii) Category Mapper, (iii) Data Explorer and (iv) R/Python Code engine. These elements allow users to define a population cohort based on specified inclusion/exclusion criteria, define key variables to examine

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 dataset 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 hospitalized) 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 (eFigure 3). This enables users to pre-define 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 dataset is provided in the Data Explorer with accompanying Dataset Specifications that describe the data structure (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 dataset 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 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: (i) *NeuroBlu*: utility functions which simplify data assembly processes using prebuilt SQL queries, (ii) *SurvBlu*: a framework for conducting time-to-event analyses, and (iii) *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 visualizing data.

Data privacy and ethics

Under HIPAA regulations, U.S. healthcare providers record and store personal healthcare information for the purposes of providing treatment, billing data, to support healthcare service

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

operations. De-identified data may be used for secondary analyses. The NeuroBlu platform has received a waiver of authorization for analysis of de-identified healthcare data from the WCG Institutional Review Board (Ref: WCG-IRB 1-1470336-1).

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 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 is encrypted with AES256.

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

Participant characteristics

Descriptive statistics for the NeuroBlu dataset were obtained on 31st July 2021. A total of 562,940 individuals (48.9% male) were present in the dataset with a mean age of 33.4 years (standard deviation: 18.4 years). The majority of the population was White (49.1%). 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 dataset, working age adults (aged 20 to 59 years) account for 62.2% and adults aged 60 years or above account for 9.1% (eTable 3/eFigure 6).

eTable 4 and Figure 3 illustrate the prevalence of mental disorders in the dataset. 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 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 & Alzheimer's disease, bipolar disorder, major depressive disorder and anxiety disorders among female patients. The constituent diagnostic codes are provided in eTable 6.

eTable 7 and eFigure 8 provide a breakdown by U.S. State of Residence. The most frequent U.S. 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 eTable 8 and eFigure 9. The greatest mean CGI-S scores were found in people with schizoaffective disorder (4.48, standard deviation: 1.42) and schizophrenia (4.44, std dev: 1.44) with the lowest mean score recorded in people with substance-use disorders (3.61, std dev: 1.80). A further analysis was conducted among patients with at least 2 recorded CGI-S scores to estimate the mean maximum and mean minimum score by diagnosis (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

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

and psychotic disorders which are typically characterized by periods of remission in between episodes.

The number of patients with at least one hospital visit (inpatient or emergency department) by diagnosis (eTable 10/eFigure 10) 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 hospitalization compared to 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 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 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.

eFigure 11 provides a distribution of the total number of clinical encounters (visits) per patient. Patients had a median of 5 visits recorded (interquartile range: 2 to 20 visits). eFigure 12 provides a distribution of the total follow-up duration available for each patient in the dataset. The median follow-up duration was 7 months (IQR: 1.3 to 24.4 months).

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

(i) Gender: 1,572 patients (0.23%)
(ii) Race: 186,517 patients (33.1%)
(iii) Psychiatric diagnosis: 66,064 patients (11.7%)
(iv) State: 21,919 patients (3.9%)
(v) NLP-derived MSE data: 132,690 patients (23.6%)

Patient and Public Involvement

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

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Findings to date

The NeuroBlu platform and MindLinc EHR dataset has been used to generate real world evidence through largescale mental health epidemiology studies as well as to develop automated NLP, data visualization and predictive analytic tools based on real world EHR data.

Epidemiology of mental disorders

Data from people with substance use disorders has been analyzed to investigate its impact among minority ethnic people,[49] and demonstrated increased rates of psychiatric hospitalization among children and adolescents[50] and increased risk of psychiatric comorbidities and hospitalization 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 has been analyzed to demonstrate that antidepressant prescribing reduced in response to U.S. 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 analyzed to generated real world evidence, the dataset has been used to develop novel data visualization and predictive analytic methods to support the analysis of mental health real world data including a data visualization 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 visualization 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 dataset, as described previously.[33,60]

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Strengths and limitations

The NeuroBlu dataset contains a rich source of de-identified EHR data that supports the generation of real-world evidence to improve understanding of mental disorders and factors affecting clinical outcomes. NeuroBlu includes data on over 560,000 people attending 25 U.S. mental healthcare providers over a period of 21 years. The large sample size and long duration of follow-up enables 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 randomized controlled trials (RCTs) or prospective observational studies (for example, comparative effectiveness studies on multiple medications or assessing the impact of substance use disorders on clinical outcomes).

NeuroBlu is built upon a robust de-identification pipeline that ensures no personal identifiable data are available to users who analyze the data. A strong data security framework that employs internationally validated encryption standards protects the anonymized dataset from being accessed outside the trusted research environment of NeuroBlu. These safeguards ensure the security of the dataset while still enabling users to easily analyze 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 visualize 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 analyze the data while ensuring the data remain secure within the trusted research environment. These features allow users from a wide range of backgrounds to analyze the data without requiring expertise in statistical software while at the same time enabling expert data scientists to perform advanced analytical procedures on the dataset using R or Python.

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

The dataset 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 dataset 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 analyzing EHR datasets 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 dataset is that clinical data may not be comprehensively completed across all domains for all patients. Clinicians typically

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

record data which 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 dataset 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 dataset 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 dataset 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 (e.g. acute physical healthcare providers or diagnoses and medications recorded in community primary care settings) are not available within the NeuroBlu dataset.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Future plans

As patients continue to receive care from healthcare providers using the MindLinc EHR, the NeuroBlu dataset will grow over time increasing in sample size and follow-up duration. We plan to analyze these data to generate real-world evidence 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 personalizing 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 enables the MindLinc dataset to be analyzed 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 organizations who wish to analyze the dataset 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 de-identification pipeline and trusted research environment to support real world evidence generation from their own datasets. 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.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare real-world data analytics

Figure captions

Figure 1: NeuroBlu Cohort Builder illustrating the population characteristics of female patients diagnosed with Post Traumatic Stress Disorder who have received sertraline

Figure 2: NeuroBlu R Code Engine which 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

Figure 3: Prevalence of mental disorders in the NeuroBlu dataset

Figure 4: Mean maximum and mean minimum CGI-S score by diagnosis

Figure 5: Percentage of patients with at least one NLP-derived MSE feature by diagnosis

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Further details

Competing interests: All authors have completed the ICMJE uniform disclosure form and declare: 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.

Author contributions: 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, MR, JS and SK) contributed to cohort profile design, manuscript preparation and approved the final version.

Ethics approval: The NeuroBlu platform has received a waiver of authorization for analysis of de-identified healthcare data from the WCG Institutional Review Board (Ref: WCG-IRB 1-1470336-1).

Source of funding: The NeuroBlu dataset and MindLinc EHR system are owned and funded in full by Holmusk.

Data availability statement: All data relevant to the study are included in the article or uploaded as supplementary information.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

References

1 Rehm J, Shield KD. Global Burden of Disease and the Impact of Mental and Addictive Disorders. *Curr Psychiatry Rep* 2019;**21**:10. doi:10.1007/s11920-019-0997-0

2 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. https://www.samhsa.gov/data/sites/default/files/reports/rpt29393/2019NSDUHFFRPDFWHTML/2019NSDUHFFR1PDFW090120.pdf

3 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.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1563985/

4 Seabury SA, Axeen 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. doi:10.1377/hlthaff.2018.05246

5 Degenhardt L, Charlson F, Ferrari A, *et al.* 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. *The Lancet Psychiatry* 2018;**5**:987–1012. doi:10.1016/S2215-0366(18)30337-7

6 Mental Health America. 2020 Access to Care Data. Ment. Heal. Am. 2020.https://www.mhanational.org/issues/2020/mental-health-america-access-care-data#four

7 Maissenhaelter BE, Woolmore AL, Schlag PM. Real-world evidence research based on big data: Motivation—challenges—success factors. *Onkologe* 2018;**24**:91–8. doi:10.1007/s00761-018-0358-3

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. doi:10.1016/j.drudis.2017.12.002

9 Corrigan-Curay J, Sacks L, Woodcock J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. *JAMA - J Am Med Assoc* 2018;**320**:867–8. doi:10.1001/jama.2018.10136

10 Klonoff DC. The New FDA Real-World Evidence Program to Support Development of Drugs and Biologics. *J Diabetes Sci Technol* 2020;**14**:345–9. doi:10.1177/1932296819832661

11 Roberts MH, Ferguson GT. Real-World Evidence: Bridging Gaps in Evidence to Guide Payer Decisions. *PharmacoEconomics - Open* 2021;**5**:3–11. doi:10.1007/s41669-020-00221-y

12 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–e1912869. doi:10.1001/jamanetworkopen.2019.12869

13 Johnson AEW, Pollard TJ, Shen L, *et al.* MIMIC-III, a freely accessible critical care database. *Sci Data* 2016;**3**:160035. doi:10.1038/sdata.2016.35

14 OMOP Common Data Model – OHDSI. 2022.https://www.ohdsi.org/data-standardization/the-common-data-model/

15 NORC Data Enclave | NORC.org. norc. 2022.https://www.norc.org/Research/Capabilities/Pages/data-enclave.aspx

16 Administrative Data Research Facility (ADRF) – Coleridge Initiative. 2022.https://coleridgeinitiative.org/adrf/

17 Stewart R, Davis K. 'Big data' in mental health research: current status and emerging possibilities. *Soc Psychiatry Psychiatr Epidemiol* 2016;**51**:1055–72. doi:10.1007/s00127-016-1266-8

18 Peterson BS. Editorial: Biomarkers in precision medicine for mental illnesses. *J Child Psychol Psychiatry Allied Discip* 2020;**61**:1279–81. doi:10.1111/jcpp.13357

19 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**. doi:10.1136/bmjopen-2015-008721

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare real-world data analytics

20 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. doi:10.1371/journal.pone.0190703

21 Cardinal RN. Clinical records anonymisation and text extraction (CRATE): An open-source software system. *BMC Med Inform Decis Mak* 2017;**17**:50. doi:10.1186/s12911-017-0437-1

22 Smoller JW. The use of electronic health records for psychiatric phenotyping and genomics. *Am J Med Genet Part B Neuropsychiatr Genet* 2018;**177**:601–12. doi:10.1002/ajmg.b.32548

23 Holderness E, Miller N, Cawkwell P, *et al.* Analysis of risk factor domains in psychosis patient health records. *J Biomed Semantics* 2019;**10**:19. doi:10.1186/s13326-019-0210-8

24 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* Published Online First: 2020. doi:https://doi.org/10.1016/j.schres.2020.09.016

25 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. doi:10.1016/j.schres.2020.05.007

26 Raket LL, Jaskolowski J, Kinon BJ, *et al.* Dynamic ElecTronic hEalth reCord deTection (DETECT) of individuals at risk of a first episode of psychosis: a case-control development and validation study. *Lancet Digit Heal* 2020;**2**:e229–39. doi:10.1016/S2589-7500(20)30024-8

27 Jackson RG, Patel R, Jayatilleke 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**. doi:10.1136/bmjopen-2016-012012

28 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. doi:10.1136/bmjopen-2020-042949

29 Holderness E, Cawkwell P, Bolton K, *et al.* 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. doi:10.18653/v1/w19-1915

30 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 Inf* 2020;**8**:e17376. doi:10.2196/17376

31 Ford T, Mansfield KL, Markham S, *et al.* The challenges and opportunities of mental health data sharing in the UK. *Lancet Digit Heal* 2021;**3**:e333–6. doi:10.1016/s2589-7500(21)00078-9

32 Holmusk. NeuroBlu: The Evidence Footprint. Building the future of behavioral health with Real-World Data and analytics. 2021.https://www.neuroblu.ai/

33 Mukherjee SS, Yu J, Won Y, *et al.* Natural Language Processing-Based Quantication of the Mental State of Psychiatric Patients. *Comput Psychiatry* 2020;**4**:76. doi:10.1162/cpsy_a_00030

34 Gersing K, Krishnan R. Clinical Computing: Clinical Management Research Information System (CRIS). *Psychiatr Serv* 2003;**54**:1199–200. doi:10.1176/appi.ps.54.9.1199

35 Response to the Goldacre Review - "Trusted Research Environments and Data Management". HDR UK. 2021.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* 2018;**2017**:1044–50.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977668/

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.https://doi.org/10.1007/978-3-030-70558-9_13

38 Patel R, Jayatilleke N, Broadbent M, *et al.* Negative symptoms in schizophrenia: a study in a large

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

clinical sample of patients using a novel automated method. *BMJ Open* 2015;**5**. doi:10.1136/bmjopen-2015-007619

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. doi:10.1093/schbul/sbx197

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**. doi:10.1136/bmjopen-2014-007504

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. doi:DOI: 10.1192/j.eurpsy.2020.39

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* Published Online First: 7 October 2020. doi:10.1093/schbul/sbaa126

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**:1178222618792860–1178222618792860. doi:10.1177/1178222618792860

44 Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. *Psychiatry (Edgmont)* 2007;**4**:28–37.https://pubmed.ncbi.nlm.nih.gov/20526405

45 *Diagnostic and statistical manual of mental disorders : DSM-IV*. Fourth edition. Washington, DC : American Psychiatric Association, [1994] ©1994 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 J Ment Sci* 1979;**134**:382–9. doi:10.1192/bjp.134.4.382

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**:38–52. doi:10.1017/S1092852914000601

49 Wu LT, 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. doi:10.1016/j.jpsychires.2013.08.022

50 Wu LT, 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. doi:10.1016/j.jpsychires.2011.06.012

51 Wu LT, 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. doi:10.1016/j.jpsychires.2012.12.009

52 Lee TS, Assam PN, Gersing KR, *et al.* The effectiveness of antidepressant monotherapy in a naturalistic outpatient setting. *Prim Care Companion J Clin Psychiatry* 2012;**14**. doi:10.4088/PCC.12m01364

53 Gersing KR, Sheehan JJ, Burchett B, *et al.* Use of augmentation agents for treating depression: Analysis of a psychiatric electronicmedical record data set. *Psychiatr Serv* 2014;**65**:1062–5. doi:10.1176/appi.ps.201300288

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. doi:10.1007/s11126-015-9389-8

55 Anisha Balani Sheryl Ker, Jennifer Hsu, Dustin Demoss, Miguel E. Rentería Sankha S. Mukherjee AJR. T4. Buprenorphine treatment retention in individuals with opioid use disorder: insights from real-world dataset. In: *American Society of Clinical Psychopharmacology*. 2021. https://pmg.joynadmin.org/documents/1005/60ad2955ef0d26cf5e684539.pdf#page=181

56 Mane KK, Owen P, Schmitt C, *et al.* Visual analytics to optimize patient-population evidence delivery

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

for personalized care. In: *2013 ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics, ACM-BCB 2013*. Association for Computing Machinery 2013. 484–8. doi:10.1145/2506583.2506608

57 Akushevich I, Kravchenko J, Gersing K, *et al.* 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. doi:10.1145/2506583.2506618

58 Pan JJ, 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. doi:10.1177/1740774508100983

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. doi:10.1016/j.jbi.2011.09.003

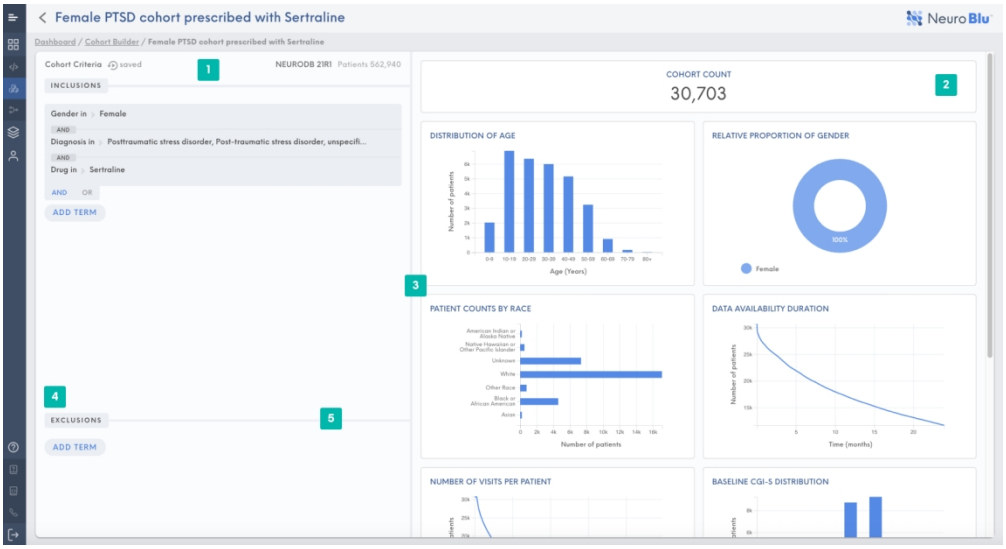
60 Sankha Mukherjee Yida Won, Mary McClay, Lu Wang, John Rush, Joydeep Sarkar JW. W35. Natural language processing-based quantification of the mental state of psychiatric patients. In: *American Society of Clinical Psychopharmacology*. 2021. https://pmg.joynadmin.org/documents/1005/60ad2955ef0d26cf5e684539.pdf#page=151

61 Hahn T, Nierenberg AA, Whitfield-Gabrieli S. Predictive analytics in mental health: applications, guidelines, challenges and perspectives. *Mol Psychiatry* 2017;**22**:37–43. doi:10.1038/mp.2016.201

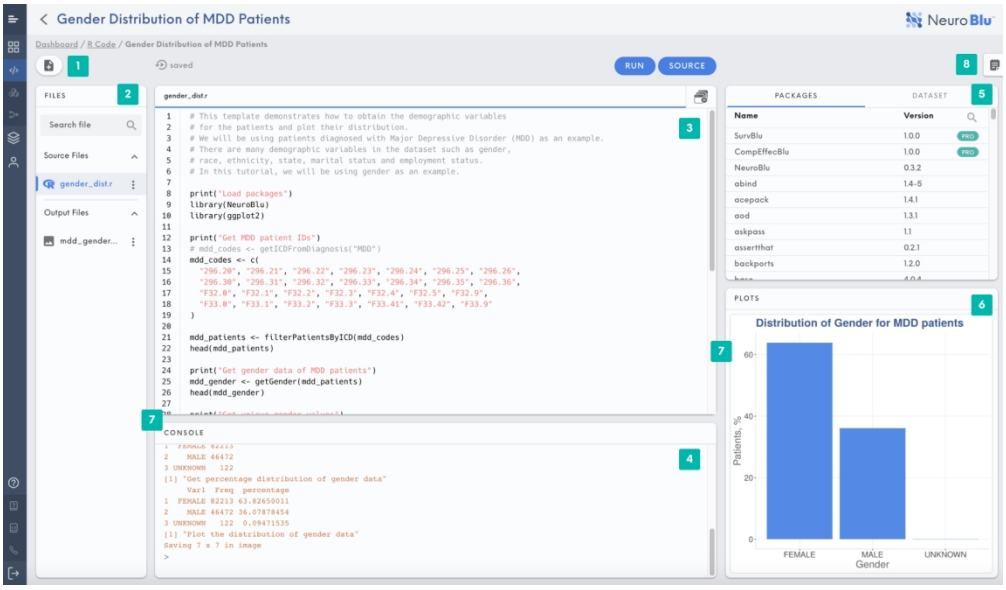
62 Holmusk. NeuroBlu Award 2021. 2021.https://www.holmusk.com/news/neuroblu-award

For peer review only

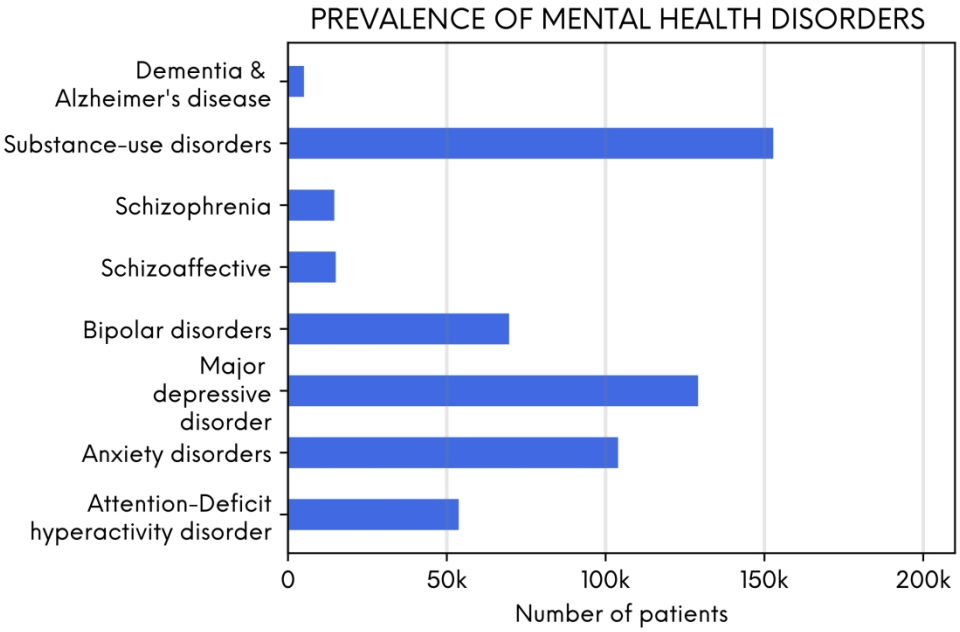
For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml



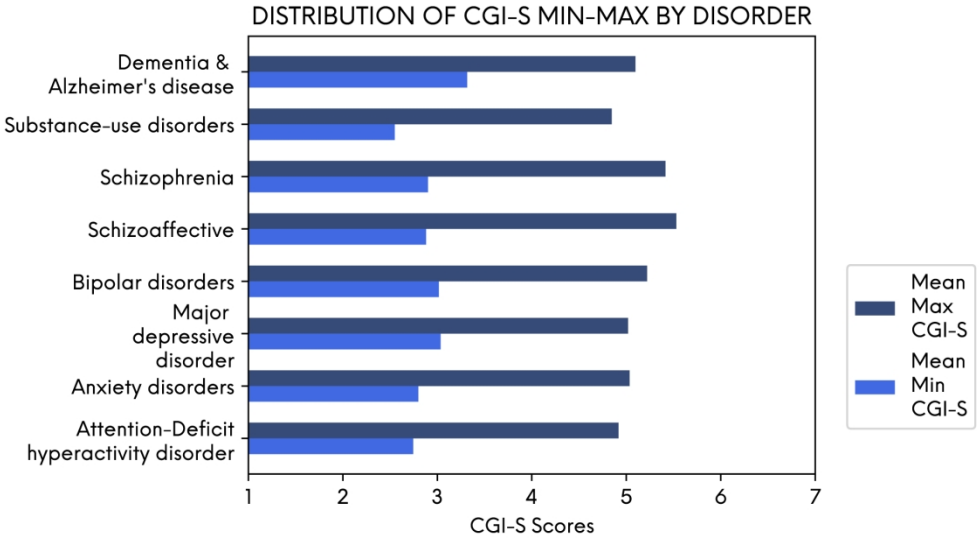
NeuroBlu Cohort Builder illustrating the population characteristics of female patients diagnosed with Post Traumatic Stress Disorder who have received sertraline



NeuroBlu R Code Engine which 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

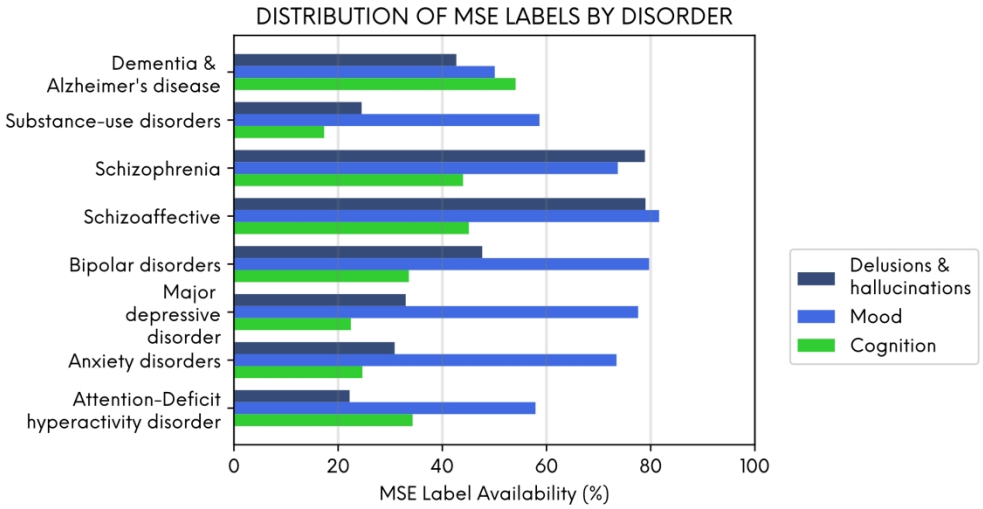


Prevalence of mental disorders in the NeuroBlu dataset



Mean maximum and mean minimum CGI-S score by diagnosis

177x99mm (300 x 300 DPI)



Percentage of patients with at least one NLP-derived MSE feature by diagnosis

Online supplementary material

Cohort profile: NeuroBlu, an Electronic Health Record Trusted Research Environment to support mental healthcare real-world data analytics

Table of contents

Table of contents 1

1. NeuroBlu Common Data Model (CDM) pipeline 3

(i) Structural Normalization 3

(ii) Semantic Normalization 4

(iii) Data Validation 4

2. Data variables 5

(i) Clinical Data Tables 5

person 6

observation_period 8

visit_occurrence 9

condition_occurrence 10

drug_exposure 11

measurement 12

observation 14

(ii) Family History 17

(iii) Derived Measurements 18

(iv) Mental State Examination 23

eTable 1: Description of U.S. clinics contributing EHR data to NeuroBlu 31

eTable 2: NeuroBlu dataset race distribution 32

eTable 3: NeuroBlu dataset age distribution 32

eTable 4: NeuroBlu dataset diagnosis distribution 32

eTable 5: Gender distribution by diagnosis 33

eTable 6: ICD-9/ICD-10 code definitions for diagnostic category 34

eTable 7: U.S. State of Residence 65

eTable 8: Mean CGI-S score by diagnosis 67

eTable 9: Mean maximum and minimum CGI-S score by diagnosis 68

eTable 10: Number of patients with at least one hospital visit by diagnosis 69

eTable 11: SAV features included in MSE NLP analysis (eTable 12) 70

eTable 12: Number of patients with at least one MSE feature by diagnosis 71

eFigure 1: MindLinc EHR data de-identification and harmonization pipeline 72

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

eFigure 2: OMOP CDM schematic 73

eFigure 3: NeuroBlu Category Mapper illustrating the classification of individual drugs into different types of medication 74

eFigure 4: NeuroBlu Data Explorer which provides a summary of the dataset and allows the top 50 rows of each table to be viewed 75

eFigure 5: NeuroBlu dataset race distribution 76

eFigure 6: NeuroBlu dataset age distribution 77

eFigure 7: Gender distribution by diagnosis 78

eFigure 8: U.S. State of Residence 79

eFigure 9: Mean CGI-S score by diagnosis 80

eFigure 10: Percentage of patients with at least one hospital visit by diagnosis 81

eFigure 11: Distribution of number of visits per patient 82

eFigure 12: Distribution of duration of follow-up 83

For peer review only

Online supplementary material

1. NeuroBlu Common Data Model (CDM) pipeline

The NeuroBlu Data Model is based on the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) developed by the Observation Health Data Sciences and Informatics (OHDSI) community (eFigure 1). The model enables data contained within disparate observational databases to be transformed into a common format (the "data model") as well as a common representation (terminologies, vocabularies, coding schemes). This allows systematic analyses to be performed using a library of standard analytical routines that have been written based on the OMOP CDM.

The NeuroBlu Data Model includes the following modifications of the OMOP CDM to better support the representation of mental healthcare data:

1. Representing ICD9/ICD10 diagnostic codes in the CONDITION_OCCURRENCE table rather than the specified SNOMED concepts, as users typically have more familiarity with the ICD code sets than the SNOMED code set.
2. Adding the current value of mutable demographic attributes such as marital status, education, and occupation to the PERSON to allow users to access this information without needing to join a separate table.
3. Adding custom tables to support derived data and other data elements that do not have a natural home in the OMOP CDM. This includes derived measurements like smoothened CGI-S scores, imputed MADRS scores, and Mental State Examination (MSE) data obtained through natural language processing (NLP).

Once the MindLinc data are de-identified, transferred, and loaded into the relational PostgreSQL database, the remaining data integration processes required to ensure the data are ready for analysis within the NeuroBlu dataset include structural normalization, semantic normalization, and data validation.

(i) Structural Normalization

Structural normalization is the process of transforming the data into a single schema or layout. For example, to examine patient demographic information in a consistent way, regardless of how it was stored in the originating EHRs, the various patient and member demographic data are organized into a single, commonly structured PERSON table. This requires establishing and implementing various transformation rules that get applied to the data, such as those described below:

1. **Source-to-Target Mapping Logic:** this refers to the conveyance of information from a specific place in the source data to its related place in the target data model. For example, the MindLinc EHR source data contains a BACKGROUND table to record patient demographic attributes, such as RACE and SEX. In order for these attributes to end up in the appropriate place in the NeuroBlu Data Model, they need to be mapped from the relevant source field to the associated target field. In this example, the information in RACE and SEX MindLinc database would end up in the RACE_SOURCE_VALUE and GENDER_SOURCE_VALUE fields of the PERSON table in the NeuroBlu database.
2. **Inclusion/Exclusion Logic:** this refers to the criteria applied to the data in order for them to be included in the NeuroBlu Data Model. For example, information from the MindLinc source data that is marked as "entered in error" or "erroneous" brings into question whether it is truly evidence from real world data and in turn is excluded from the NeuroBlu Data Model.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

3. **De-duplication Logic:** this refers to the removal of duplicative or outdated information, such that only up-to-date data are pulled into the NeuroBlu Data Model and are only pulled in once. Depending on the source data elements of interest and the mechanism by which updates are applied to them, the method for de-duplication will vary. In the case of the MindLinc data, a destructive refresh approach is taken across all data tables such that the most recently acquired data will always replace the older data.

(ii) Semantic Normalization

Semantic normalization is the process of translating the data to a single set of units and vocabulary. For example, in order to specifically examine female patients in the aforementioned PERSON table, the various ways the patient and member demographic data codify the concept of "female" need to be examined such that a translation between the source codes (e.g. "F", "FEM", "FEMALE", "0", "396123", etc.) and the NeuroBlu standard concept ("FEMALE") can be created and applied to the data. These translations are generated through a combination of direct-mapping, text-string-matching, and NLP based on a thorough examination of the native codes, their descriptions, any supporting information provided about them (e.g. reference ranges for scores, generic name for medications, etc.), and, if necessary, expert input from the healthcare provider organization they are originating from.

A number of industry standard reference code sets are leveraged as part of the semantic normalization process for the MindLinc data, including RxNorm for drug mapping and ICD9/ICD10 for diagnostic code mapping. The target concept sets for all normalized data elements are the Standardized Vocabulary provided as part of the OMOP CDM standard.

(iii) Data Validation

A core component of data acquisition and integration is ensuring that the resultant data set is of the highest quality possible, where quality is defined as a function of conformance, completeness, correctness, and consistency. In order to measure the quality of newly integrated data against these key dimensions, the NeuroBlu CDM pipeline employs an approach to data validation that uses the wealth of data available during the process to produce a composite measure of what certain key data points typically look like in different situations, and then evaluates the data against these expectations. Doing this efficiently requires validation of the data as they flow through the processing pipeline, which in turn necessitates a robust strategy for determining key quality indication metrics, generating those metrics in a systematic manner, and distilling those metrics into meaningful insights.

The Data Validation process works in concert with the above normalization processes to identify issues or inconsistencies in the data that need to be addressed and provide guidelines on how those various issues or inconsistencies should be handled. Erroneous data that can't be fixed at the source are either remedied or removed by NeuroBlu's Extract, Transform, Load (ETL) processes, while quirks or unexpected features of the data that are not necessarily errors are systematically logged.

The end result of these data integration processes is the creation of a single normalized and validated Common Data Model for use in the NeuroBlu trusted research environment.

For peer review only

Online supplementary material

2. Data variables

(i) Clinical Data Tables

The clinical data tables contain the core information about the clinical events that occurred longitudinally during valid Observation Periods for each person, as well as demographic information for the person. In addition to the standardized clinical data tables defined in the OMOP CDM, the NeuroBlu dataset also contains other tables that are not defined in the OMOP CDM including data on Family History, Derived Measurements and Mental State Examination and which are described in greater detail in sections (ii), (iii) and (iv). The tables form part of a relational database in which "_id" columns may be joined between tables to assemble clinical data at individual person level.

Table name	Type	Content
person	CDM	Socio-demographic information of each patient including year of birth, gender, race, location of residence, occupational status and education status.
observation_period	CDM	Time interval for which a patient has data recorded. Data from this table is used to filter cohorts according to minimum duration within the dataset to support longitudinal analyses.
visit_occurrence	CDM	Records of each visit and its type. Data may be obtained from this table on the number of contacts with the treating mental healthcare service and the type of contact (e.g. outpatient or inpatient).
condition_occurrence	CDM	Records of clinical diagnosis in the form of ICD-9-CM and ICD-10-CM codes. Data from this table is used to generate population cohorts based on primary mental disorder diagnosis and to obtain data on comorbid diagnoses.
drug_exposure	CDM	Records of any exposure to prescribed medications including the names of prescribed medications, start and stop dates, dose and frequency of administration.
measurement	CDM	Records of direct clinical measurements using clinician-rated structured rating scales.
observation	CDM	NLP-derived records of social factors which could contribute towards risk of developing a mental disorder and/or poorer clinical outcomes. This table can also be used to obtain social factor data as outcomes (e.g. social and occupational functioning)
family_history	Non-CDM	Provides recorded data on family history of mental disorders
derived_measurement	Non-CDM	Clinical measurements which have been transformed or derived from recorded CGI-S scores.
derived_mse	Non-CDM	NLP-derived mental state examination (MSE) data.
CDM: OMOP Common Data Model Non-CDM: NeuroBlu customized data structure (not defined in OMOP)		

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

person

The person table contains records that uniquely identify each person in the dataset. This table also contains demographic information about each individual (excluding data removed by the de-identification process). All other clinical data tables refer to this table using the person_id field which is the unique record identifier for each individual represented in the dataset.

Columns

Field	Type	Description
person_id	int8	A unique identifier for each person. This value is used to join onto other tables in NeuroBlu to obtain clinical data at individual person level.
gender_concept_id	int4	A foreign key that refers to an identifier in the concept table for the unique gender of the person, belonging to the 'Gender' Vocabulary. Gender data are obtained by joining on this value (see below).
year_of_birth	int4	The year of birth of the person.
race_concept_id	int4	A foreign key that refers to an identifier in the concept table for the unique race of the person, belonging to the 'Race' Vocabulary. Race data are obtained by joining on this value (see below).
ethnicity_concept_id	int4	A foreign key that refers to an identifier in the concept table for the unique ethnicity of the person, belonging to the 'Ethnicity' Vocabulary. Ethnicity data are obtained by joining on this value (see below).
zip_code_source_value	varchar	3-digit ZIP code of the person's location.
city	varchar	City name based on 3-digit ZIP code.
state	varchar	State name based on 3-digit ZIP code.
marital_status	varchar	Marital status of the person.
occupation	varchar	Employment status of the person.
education	int2	Number of years of formal education. When education_source_value is not available, defaults to 0
education_source_value	varchar	Source value for education. Not always available.

Conventions

No.	Convention Description
1	All tables representing patient-related domains have a foreign-key reference to the person_id field in the person table.
2	Each person record has associated demographic attributes which are assumed to be constant for the patient throughout the course of their periods of observation. For example, the location or gender is expected to have a unique value per person, even though in life these data may change over time.
3	The month and day of birth for a person is not recorded. Therefore, value defaults to Jan 1st.
4	Ethnicity and Race in the OMOP CDM follow the 1997 Office of Management and Budget (OMB) standards for data on race and ethnicity.

For peer review only

Online supplementary material

Gender Vocabulary

concept_id	concept_name
8507	MALE
8532	FEMALE
8551	UNKNOWN

Race Vocabulary

concept_id	concept_name
8515	Asian
8516	Black or African American
8522	Other Race
8527	White
8552	Unknown
8557	Native Hawaiian or Other Pacific Islander
8657	American Indian or Alaska Native

Ethnicity Vocabulary

concept_id	concept_name
759814	Ethnic group unknown
38003563	Hispanic or Latino
38003564	Not Hispanic or Latino

Marital Status Vocabulary

marital_status
death of spouse
divorced
engaged
married
separated
single
together
unknown

Occupation Vocabulary

occupation
Disabled
Employed
Retired
Student
Unemployed
Unknown

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

observation_period

The observation_period table contains records which uniquely define the spans of time for which a patient is expected to have clinical events recorded within the source systems, even if no events in fact are recorded (healthy patient with no healthcare interactions).

Columns

Field	Type	Description
observation_period_id	serial	A unique identifier for each observation period.
person_id	int8	A foreign key identifier to the person for whom the observation period is defined. The demographic details of that person are stored in the person table.
observation_period_start_date	date	The start date of the observation period for which data are available from the data source.
observation_period_end_date	date	The end date of the observation period for which data are available from the data source.
period_type_concept_id	int4	A foreign key identifier to the predefined Concept in the Standardized Vocabularies reflecting the source of the observation period information, belonging to the 'Obs Period Type' Vocabulary.

Conventions

No.	Convention Description
1	Each person has to have at least one observation period.
2	One person may have one or more disjoint observation periods, during which times analyses may assume that clinical events would be captured if observed.
3	Each person can have more than one valid observation period record, but no two observation periods can overlap in time for a given person.
4	As a general assumption, during an observation period any clinical event that happens to the patient is expected to be recorded. Conversely, the absence of data indicates that no clinical events occurred to the patient.
5	Both the start date and the end date of the clinical event has to be between observation_period_start_date and observation_period_end_date.
6	For EHR data, the observation period cannot be determined explicitly, because patients usually do not announce their departure from a certain healthcare provider. The extract, transform, load (ETL) process will have to apply some heuristic to make a reasonable estimate on what the observation_period should be.

For peer review only

Online supplementary material

visit_occurrence

The visit_occurrence table contains the spans of time a patient continuously receives medical services from one or more providers at a care site in a given setting within the health care system.

Columns

Field	Type	Description
visit_occurrence_id	int8	A unique identifier for each person's visit or encounter at a healthcare provider.
person_id	int8	A foreign key identifier to the person for whom the visit is recorded. The demographic details of that person are stored in the person table.
visit_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies belonging to the 'Visit' Domain.
visit_start_datetime	timestamp	The date and time when the visit started. The time is not captured in the source and defaults to 00:00:00.
visit_end_datetime	timestamp	The date and time when the visit ended. Not recorded in the source. Takes the same value as visit_start_datetime.
visit_type_concept_id	int4	A foreign key identifier to the predefined Concept in the Standardized Vocabularies reflecting the source data from which the visit record is derived from, belonging to the 'Visit Type' Vocabulary. Visit type data are obtained by joining on this value (see below).
provider_id	int8	A foreign key to the primary care provider the person is seeing.
care_site_id	int8	A foreign key to the site of primary care.
visit_source_value	varchar	Source value for visit_concept_id.

Conventions

No.	Convention Description
1	A visit occurrence is recorded for each engagement of a patient with the healthcare system. It reflects any interaction with the healthcare service, not just in person clinical interactions.
2	Valid Visit Concepts belong to the 'Visit' domain.
3	At any one day, there could be more than one visit.
4	Visit times and visit end dates are not captured in the source. They resort to 00:00:00 and visit start dates, respectively.

Visit types

The table below describes the different visit types documented in the visit_occurrence table.

concept_id	concept_name
262	Emergency Room and Inpatient Visit
9201	Inpatient Visit
9202	Outpatient Visit
32693	Health examination
581458	Pharmacy visit
581477	Office Visit
42898160	Non-hospital institution Visit

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

condition_occurrence

The condition_occurrence table contains the presence of a disease or medical condition recorded as a diagnosis according to the International Classification of Disease (ICD), which is either observed by a provider or reported by the patient. In the NeuroBlu dataset, diagnoses are recorded as ICD-9-CM and ICD-10-CM codes in the source.

Columns

Field	Type	Description
condition_occurrence_id	serial	A unique identifier for each condition occurrence event.
person_id	int8	A foreign key identifier to the person for whom the condition is recorded. The demographic details of that person are stored in the person table.
condition_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies belonging to the 'Condition' Domain. Joining on this value extracts the corresponding ICD-9-CM or ICD-10-CM diagnostic code.
condition_start_datetime	timestamp	The date and time when the instance of the condition is recorded. The time is not captured in the source and defaults to 00:00:00.
provider_id	int8	A foreign key to the primary care provider the person is seeing.
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the condition was diagnosed.
condition_source_value	varchar(255)	Source value for condition_concept_id.

Conventions

No.	Convention Description
1	Past diagnoses ('history of') are not recorded in the condition_occurrence table. Instead, they are listed in the observation table.
2	The exact start date and end date for a condition is not recorded explicitly. Rather, the start date contains the date when the diagnosis was made, and the end date is left empty.

For peer review only

Online supplementary material

drug_exposure

The drug_exposure table contains records about the utilization of a prescribed medication when ingested or introduced to the body. Prescribed medications are inferred from clinical events associated with orders, prescriptions written, pharmacy dispensing, procedural administrations, and other patient-reported information.

Columns

Field	Type	Description
drug_exposure_id	serial	A unique identifier for each condition exposure event.
person_id	int8	A foreign key identifier to the person for who is subjected to the drug. The demographic details of that person are stored in the person table.
drug_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies belonging to the 'Drug' Domain. Joining on this value extracts the name of the prescribed medication.
drug_exposure_start_datetime	timestamp	The start date and time for the current instance of drug utilization. Valid entries include a start date of a prescription or the date on which a drug was administered. The time is not captured in the source and defaults to 00:00:00.
drug_exposure_end_datetime	timestamp	The end date and time for the current instance of drug utilization. It could be a known or an inferred date and denotes the last day at which the patient was still exposed to the drug during an instance of drug utilization.. The time is not captured in the source and defaults to 00:00:00.
quantity	numeric	Total quantity of drug exposed, in mg. Not always available.
days_supply	int4	The number of days of drug exposure. Not always available.
provider_id	int8	A foreign key to the primary care provider the person is seeing.
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the drug exposure was initiated.
drug_source_value	varchar(50)	Source value for drug_concept_id.

Conventions

No.	Convention Description
1	Valid Concepts for the drug_concept_id field belong to the 'Drug' domain. They are based on Ingredient Concept Class within the RxNorm Vocabulary.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

measurement

The measurement table contains records of structured values (numerical or categorical) obtained through systematic and standardized examination or testing of a person. In the NeuroBlu dataset, the table contains measurements of:

1. Clinical Global Impression - Severity (CGI-S)
2. Clinical Global Impression - Improvement (CGI-I)
3. Global Assessment of Functioning (GAF)
4. Montgomery-Asberg Depression Rating Scale (MADRS)

Columns

Field	Type	Description
measurement_id	serial	A unique identifier for each measurement.
person_id	int8	A foreign key identifier to the person for whom the measurement is recorded. The demographic details of that person are stored in the person table.
measurement_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies. Joining on this value specified which clinical measurement is being recorded.
measurement_datetime	timestamp	The date and time of the measurement. The time is not captured in the source and defaults to 00: 00: 00.
measurement_type_concept_id	int4	A foreign key to the predefined Concept in the Standardized Vocabularies reflecting the provenance from where the Measurement record was recorded. These belong to the 'Meas Type' vocabulary.
operator_concept_id	int4	A foreign key identifier to the predefined Concept in the Standardized Vocabularies reflecting the mathematical operator that is applied to the value_as_number. Operators are <, < = , = , > = , > and these concepts belong to the 'Meas Value Operator' domain.
value_as_number	numeric	A measurement result where the result is expressed as a numeric value.
provider_id	int8	A foreign key to the primary care provider the person is seeing.
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the measurement was recorded.
measurement_source_value	varchar	Source value for measurement_concept_id.
measurement_source_concept_id	int4	A foreign key to a Concept in the Standardized Vocabularies that refers to the code used in the source.
unit_source_value	varchar	Source value for the unit.
value_source_value	varchar	Source value associated with value_as_number column.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

custom2_str	varchar	Custom field to indicate which measurement the record is for.

Conventions

No.	Convention Description
1	Measurements are stored as attribute value pairs, with the attribute as the measurement_concept_id and the value representing the result (value_as_number).
2	The operator_concept_id is optionally given for relative Measurements where the precise value is not available but its relation to a certain benchmarking value is available. For example, this can be used for minimal detection thresholds of a test.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

observation

The observation table contains records of clinical facts about a related to medical history, the stated need for certain treatment, social circumstances, lifestyle choices, healthcare utilization patterns, etc. These records are obtained during examination or procedure and differ from Measurements in that they do not require a standardized test to generate clinical fact.

In NeuroBlu, the observation table contains the social history in the form of external environmental stressors.

Columns

Field	Type	Description
observation_period_id	serial	A unique identifier for each observation.
person_id	int8	A foreign key identifier to the person for whom the observation period is defined. The demographic details of that person are stored in the person table.
observation_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies.
observation_datetime	timestamp	The date and time of the observation. The time is not captured in the source and defaults to 00: 00: 00.
observation_type_concept_id	int4	A foreign key to the predefined Concept in the Standardized Vocabularies reflecting the provenance from where the Observation record was recorded. These belong to the 'Observation Type' vocabulary.
value_as_string	varchar(255)	Categorical value of the Result of the Observation, if applicable and available.
value_as_concept_id	int4	A foreign key to an observation result stored as a Concept ID. This is applicable to observations where the result can be expressed as a Standard Concept from the Standardized Vocabularies (e.g., positive/negative, present/absent, low/high, etc.).
qualifier_concept_id	int4	Contains all attributes specifying the clinical fact further, such as degrees, severities, drug-drug interaction alerts
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the measurement was recorded.
observation_source_value	varchar(255)	The observation code (Eg: ICD10) as it appears in the source data.
qualifier_source_value	varchar(50)	The source value associated with a qualifier to characterize the observation.

For peer review only

Online supplementary material

Conventions

No.	Convention Description
1	Observations can be stored as attribute value pairs, with the attribute as the Observation Concept and the value representing the clinical fact.
2	Valid Observation Concepts are not enforced to be from any domain though they still should be Standard Concepts.

External Stressors data derived using NLP

In the NeuroBlu dataset, the raw data for external environmental stressors (or social history) is a mixture of structured values and custom, user-defined free text. These values need to be normalized to structured values before they are used for downstream analytics. To achieve this, we have applied Natural Language Processing (NLP) to map the unstructured free text to structured labels.

The NLP model maps the semi-structured stessor data to one of the possible 61 labels (which can be categorized into 8 categories). Each label is also associated with a polarity (positive, negative or neutral) as described in the table below.

Category	Stressor Label	Polarity
Academic	Discord with teachers and classmates	Negative
Academic	Good achievement in school	Positive
Academic	Illiteracy and low-level literacy	Negative
Academic	Suspension/expulsion	Negative
Academic	Underachievement in school	Negative
Environment	Abusive environment at home	Negative
Environment	Chaotic atmosphere / stressful home environment	Negative
Environment	Cultural conflicts/issue	Negative
Environment	Exposure to Disaster, War, or Other Hostilities	Negative
Environment	Inadequate care (history from caregiver or patient)	Negative
Environment	Migration/ Acculturation Difficulty	Negative
Environment	Problem Related to Living Alone	Negative
Environment	Stable living situation	Positive
Family	Anxiety about sick person in family, dependent relative needing care at home	Negative
Family	Bereavement	Negative
Family	Disruption of Family by Separation or Divorce	Negative
Family	Inadequate or distorted communication within family	Negative
Family	Isolated family, lack of family support	Negative
Family	Legal problem in the family	Negative
Family	Parenting situation	Negative
Family	Problems in parent-child relationships and in-laws	Negative
Family	Sibling Relational Problem	Negative
Family	Substance abuse in the family	Negative
Family	Supportive family, love and respect in the family	Positive
Financial	Extreme Poverty/ Low income	Negative
Financial	Financial stability	Positive
Financial	Inadequate health/social Insurance or Welfare Support	Negative
Financial	Poor money management / no savings	Negative
Financial	Problems related to loan / creditors	Negative
Housing	Discord With Neighbor	Negative

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Housing	Homelessness	Negative
Housing	Inadequate housing	Negative
Housing	Isolated dwelling (Lack of transportation)	Negative
Housing	Unstable housing arrangement	Negative
Legal	Arrest, Imprisonment and other incarceration	Negative
Legal	Child custody or support proceedings	Negative
Legal	Litigation / legal proceedings	Negative
Legal	Offense / charges	Negative
Legal	Problem with legal personnel	Negative
Legal	Restrain order	Negative
Occupational	Change of job	Neutral
Occupational	Fired	Negative
Occupational	Job dissatisfaction	Negative
Occupational	Job satisfaction	Positive
Occupational	Occupational stress	Negative
Occupational	Threat of job loss, Underperform at work	Negative
Occupational	Unemployment	Negative
Occupational	Unstable job (changing jobs often)	Negative
Personal	Emotional abuse	Negative
Personal	Involuntary treatment	Negative
Personal	Major life events	Negative
Personal	Physical abuse (spouse)	Negative
Personal	Relationship Distress With Spouse or Intimate Partner	Negative
Personal	Sense of meaning and support from belief	Positive
Personal	Sexual abuse (spouse)	Negative
Personal	Target of Adverse Discrimination	Negative
Personal	Victim of crime	Negative
Social	Conflictual social relationships	Negative
Social	Inadequate social support	Negative
Social	Lack of social circles	Negative
Social	Supportive friends	Positive

Precision statistics for the NLP classifiers provided as the median metrics of each stressor category are provided in the table below.

Category	AUROC	Recall	Precision	F1
Academic	0.999	0.667	1.000	0.800
Environment	0.999	0.923	0.800	0.857
Family	0.984	0.519	0.759	0.619
Financial	0.998	0.500	0.833	0.667
Housing	0.993	0.417	0.633	0.473
Legal	0.985	0.708	0.813	0.745
Occupational	0.993	0.500	0.895	0.664
Social	0.990	0.500	0.833	0.625

For peer review only

Online supplementary material

(ii) Family History

The family_history table contains data on recorded diagnoses of mental disorders affecting patients' relatives in the form of ICD-10-CM codes, along with the degree of relationship relative to the patient.

Columns

Field	Type	Description
family_history_id	serial	A unique identifier for each family history record.
person_id	int8	A foreign key identifier to the person for whom the MSE is recorded. The demographic details of that person are stored in the person table.
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the data was recorded.
condition_concept_id	int4	A foreign key that refers to an identifier in the Standardized Vocabularies belonging to the 'Condition' Domain. Joining on this value extracts the corresponding ICD-9-CM or ICD-10-CM diagnostic code.
condition_source_value	varchar(255)	Source value for condition_concept_id.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

(iii) Derived Measurements

The derived_measurement table contains records of structured values (numerical or categorical) obtained through deriving recorded measurements in the measurement table. In the NeuroBlu dataset, the table contains measurements of:

1. Smoothened CGI-S
2. Derived MADRS

Columns

Field	Type	Description
derived_measurement_id	serial	A unique identifier for each derived measurement.
person_id	int8	A foreign key identifier to the person for whom the measurement is recorded. The demographic details of that person are stored in the person table.
derived_measurement_value	float4	The value of the derived measurement.
derived_measurement_date	date	The date of the derived measurement.
range_low	numeric	The lower limit of the normal range of the measurement result. The lower range is assumed to be of the same unit of measure as the measurement value.
range_high	numeric	The upper limit of the normal range of the measurement result. The upper range is assumed to be of the same unit of measure as the measurement value.
provider_id	int8	A foreign key to the primary care provider the person is seeing.
visit_occurrence_id	int8	A foreign key to the visit in the visit_occurrence table during which the measurement was recorded.
unit_source_value	varchar(50)	Source value for the unit.
custom2_str	varchar(255)	Custom field to indicate which measurement the record is for (i.e. Derived MADRS or Smoothened CGI-S)

Conventions

No.	Convention Description
1	Measurements are stored as attribute value pairs, with the attribute as the custom2_str and the value representing the result (derived_measurement_value).

For peer review only

Online supplementary material

Smoothened CGI-S

In the NeuroBlu dataset, Clinical Global Impression-Severity (CGI-S) measurements are found to be noisy, both the severity and the improvement scales. The values recorded fluctuate across measurements. Sometimes a visit doesn't have any CGI-S measurement, and sometimes multiple CGI-S measurements are recorded in a single day.

To tackle this problem, we provide smoothened values of CGI-S. Raw measurement values are smoothened and interpolated on days with visit but no CGI-S measurements (the visit is recorded in visit_occurrence, but no CGI-S is measured on that day). These smoothened and interpolated values are available in the derived_measurement table.

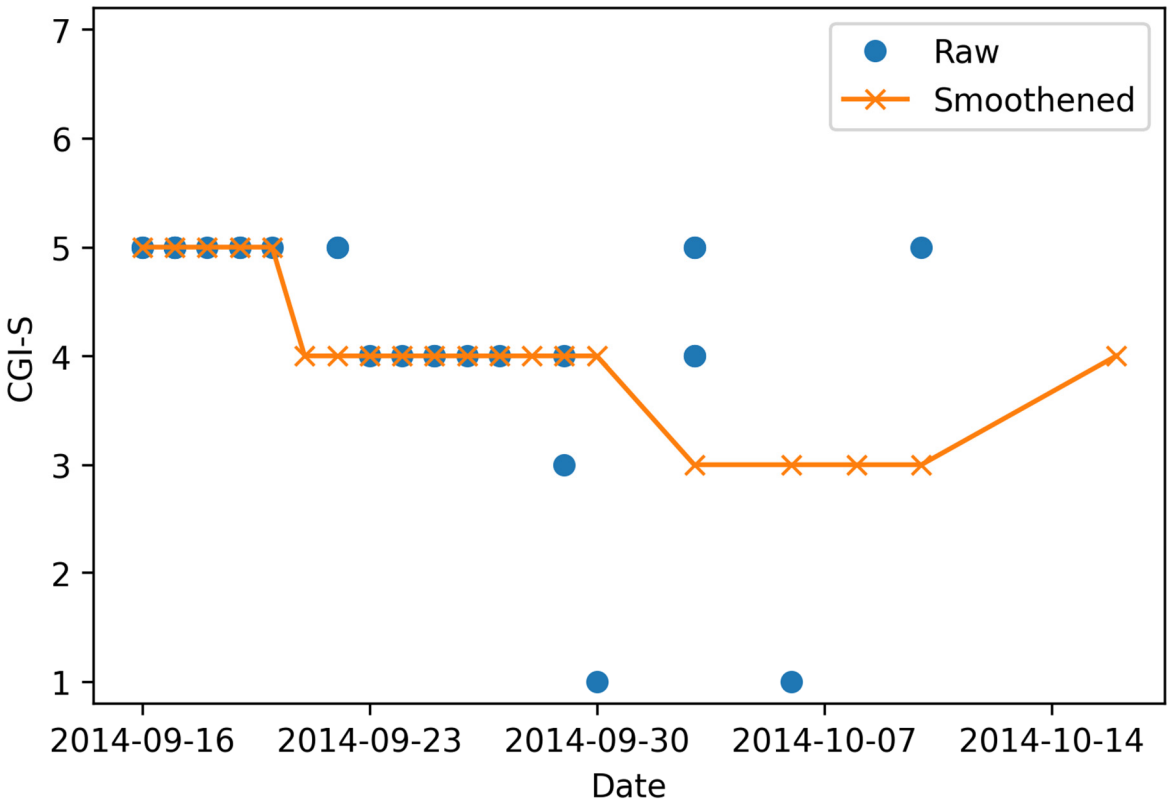
The smoothing and interpolation is done on a per-patient basis using exponential-weighted smoothing, based on the following equation:

$$s_i = \sum_{j \in J} w_{i,j} c_j$$

$$w_{i,j} = \frac{\exp(-d_{i,j}^2/(2\sigma^2))}{\sum_{j \in J} \exp(-d_{i,j}^2/(2\sigma^2))}$$

where *s_i* is the smoothened CGI-S at visit *i*, *J* is the set of all visits with CGI-S measurements, *c_j* is the measured raw CGI-S values at visit *j*, *d_i,j* is the time difference between visit *i* and *j*, in terms of days, and *σ* is the smoothing parameter, equals to 5 in our case. The equation works for days with multiple CGI measurements and days without CGI-S measurements.

The following figure shows how the smoothened CGI-S looks for an example patient.



Online supplementary material

MADRS Scores derived using CGI-S data

The Montgomery–Asberg Depression Rating Scale (MADRS) is a clinician-recorded rating scale for symptoms of depression. It comprises of 10 individual scores which are summed to yield a final score which ranges from 0 to 60. However, unlike the Clinical Global Impression-Severity (CGI-S), the NeuroBlu dataset has few documented records of MADRS scores. However, it is also one of the most widely used scores in psychiatry. We have observed that there is a fairly high correlation between the CGI-S and the final MADRS score. For this reason mapping functions have been derived that may be used to map the CGI-S score to the MADRS score.

In the event an analysis requires MADRS data, we provide values of MADRS derived from CGI-S at every timepoint with CGI-S measurement. These smoothened and interpolated values are available in the derived_measurement table.

All encounters in which both CGI-S and MADRS scores have been simultaneously measured are extracted from the dataset. In total, there are approximately 3101 such encounters. Both measurements are strongly correlated with a correlation coefficient of 0.80. This supports the possibility of mapping CGI-S to MADRS.

Upon manual inspection of the distribution of the MADRS scores for each value of CGI-S, we observed that MADRS scores have asymmetric distribution shapes for CGI-S values in the end of the spectrum, and wide distribution shapes for CGI-S values around the middle of the spectrum. For this reason, we feel that a simple linear model may not provide good confidence interval for the derived MADRS from CGI-S. We conclude that MADRS score follows the four-parameter beta distribution Beta(a, b, 0, 60), where the parameters a, b is a function of the CGI-S score, and the value is bounded between 0 and 60. Then, the values of a, b are estimated for every possible value of CGI-S using the method of moments. The derived MADRS for each CGI-S score takes value of the distribution mean, i.e. $60*a/(a+b)$. To gauge the goodness-of-fit of this approach, the R^2 value is computed and is found to be 0.66.

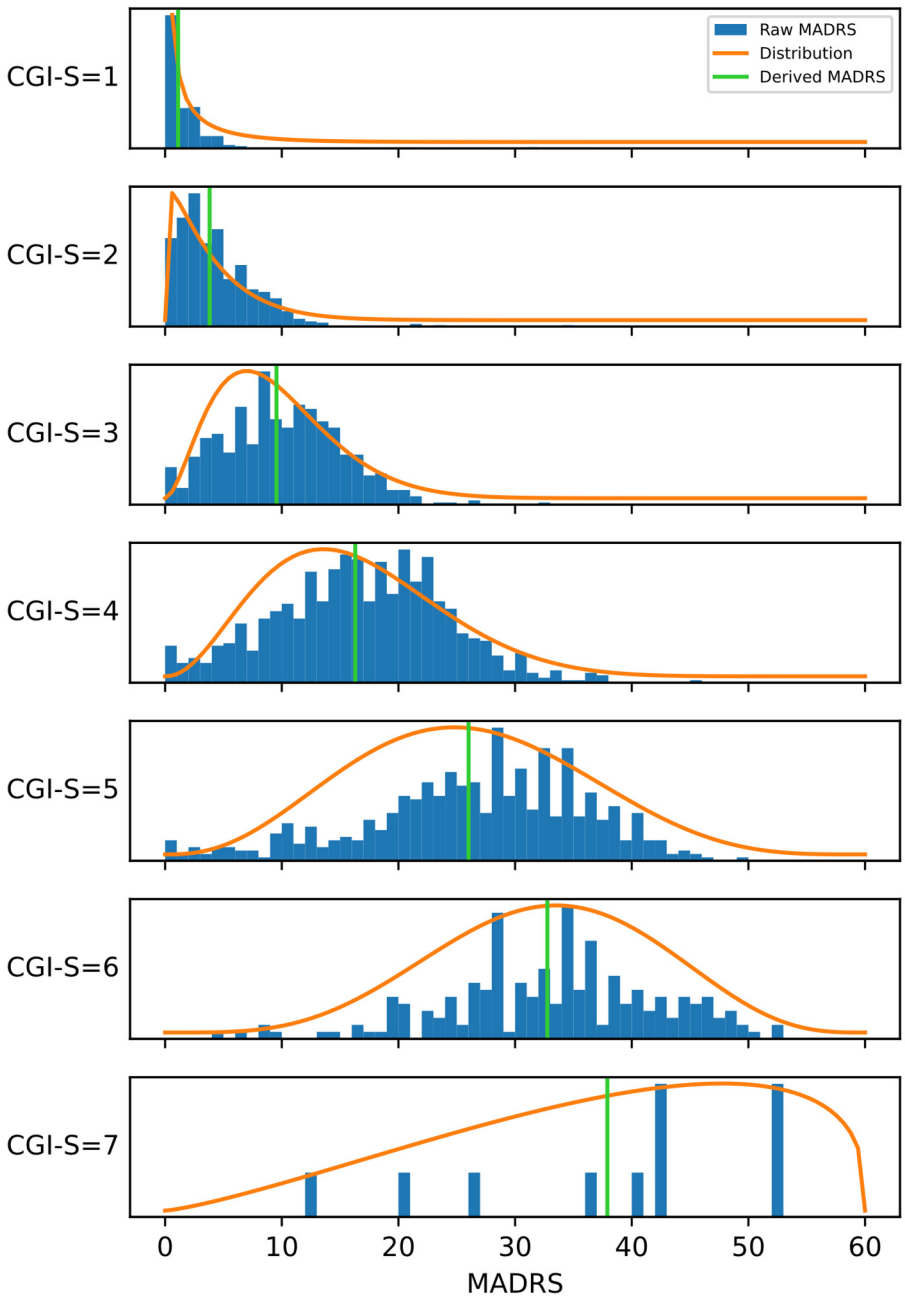
The estimated values of a and b along with the derived MADRS value for each CGI-S score are summarized in the following table.

CGI-S	a	b	Derived MADRS	95% confidence interval
1	0.1874	9.9098	1.1	0.9 to 1.3
2	1.1070	16.3081	3.8	3.6 to 4.0
3	2.8634	15.1325	9.5	9.3 to 9.8
4	3.2716	8.7725	16.3	16.0 to 16.6
5	3.7353	4.8806	26.0	25.5 to 26.6
6	5.1854	4.3100	32.8	31.9 to 33.6
7	2.2803	1.3288	37.9	32.7 to 43.1

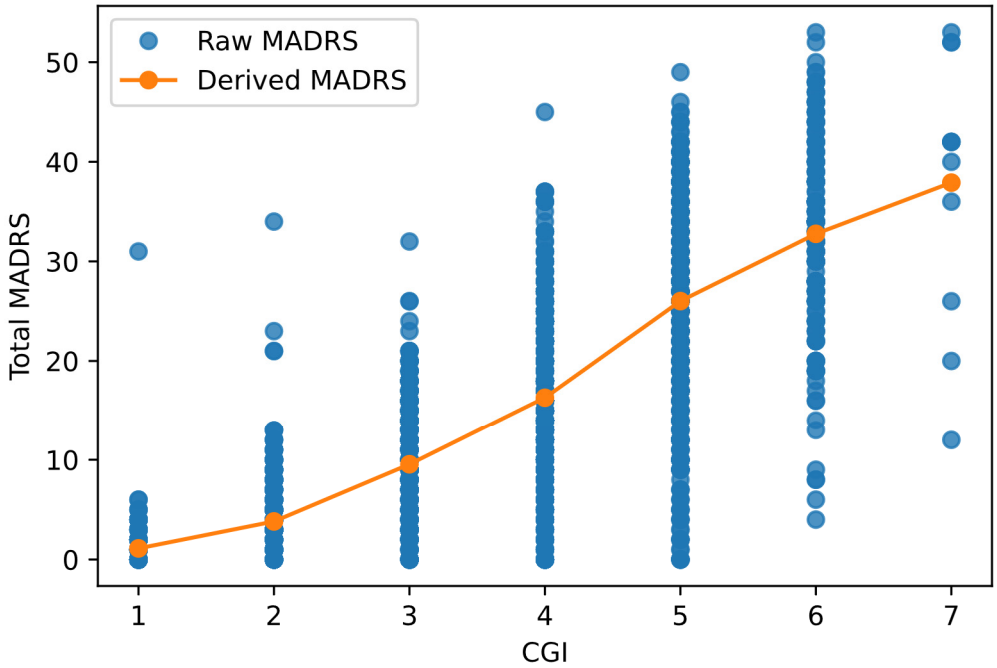
The distributions of MADRS score for each value of CGI-S are shown in the following figures. The first figure shows the histogram of MADRS scores for each value of CGI-S and the shape of the fitted beta distributions. The second figure shows the relationship between derived MADRS and CGI-S on top of a scatter plot of raw MADRS versus raw CGI-S.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material



Online supplementary material



er review only

(iv) Mental State Examination

The derived_mse table contains the cleaned form of the mental state examination (MSE) records. In the NeuroBlu dataset, the MSE data are represented as Status Assessment Vectors (SAV) which are further described subsequently.

Columns

Field	Type	Description
person_id	int8	A foreign key identifier to the person for whom the MSE is recorded. The demographic details of that person are stored in the person table.
measurement_date	date	The date of the derived MSE.
Days	int4	Number of days of measurement since the patient was first recorded in the dataset.
sav_i_cat_j	int4	i-th value of the SAV. Also correspond to the j-th value for cat category. Contains binary value. i takes value from 0 to 240.

Status Assessment Vectors

In the NeuroBlu dataset, the raw data for MSE is recorded in the form of semi-structured free text, making it difficult to analyze in their raw form.

To tackle this problem, we provide natural language processing (NLP) derived structured values of MSE. Raw MSE data for a single patient on a particular day are cleaned and passed through an NLP model that returns a 241-element vector form of MSE. This vector is called the Status Assessment Vector (SAV). SAV is a binary vector, where the value is 1 if the category-factor combination of the SAV is applicable for the patient. As an example, the 15th position in the SAV corresponds to whether a patient is responding to internal stimuli or not and value of 1 for this position means that the patient is responding to internal stimuli. Further details of the NLP pipeline are described in the following article:

Mukherjee, S.S., Yu, J., Won, Y., McClay, M.J., Wang, L., Rush, A.J. and Sarkar, J., 2020. Natural Language Processing-Based Quantification of the Mental State of Psychiatric Patients. *Computational Psychiatry*, 4, pp.76–106. DOI: http://doi.org/10.1162/cpsy_a_00030

The table below maps each element of the SAV to its location in the derived_mse table.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

SAV #	Category	Factor
0	Abnormal or psychotic thoughts	no issues
1	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (grandeur)
2	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (persecution)
3	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (religious)
4	Abnormal or psychotic thoughts	experiencing hallucinations (auditory)
5	Abnormal or psychotic thoughts	experiencing hallucinations (visual)
6	Abnormal or psychotic thoughts	experiencing hallucinations (tactile)
7	Abnormal or psychotic thoughts	History of delusions/abnormal thoughts
8	Abnormal or psychotic thoughts	History of hallucinations
9	Abnormal or psychotic thoughts	neutral/unable to categorize
10	Abnormal or psychotic thoughts	experiencing delusions (not specified)
11	Abnormal or psychotic thoughts	experiencing hallucinations (not specified)
12	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (obsessions)
13	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (paranoia)
14	Abnormal or psychotic thoughts	experiencing hallucinations (olfactory)
15	Abnormal or psychotic thoughts	responding to internal stimuli
16	Abnormal or psychotic thoughts	experiencing delusions/abnormal thoughts (sexual)
17	Affect	no issues/appropriate affect
18	Affect	neutral/unable to categorize
19	Affect	aggressive
20	Affect	reactive
21	Affect	irritable/angry
22	Affect	improved from last visit
23	Affect	issues with cognition
24	Affect	loss of energy/tiredness

For peer review only

Online supplementary material

25	Affect	blunted/restricted
26	Affect	anxious/fearful
27	Affect	tearful/depressed/sad
28	Affect	bright/positive
29	Affect	labile
30	Affect	intense
31	Affect	expansive/overly animated
32	Affect	generally odd or out of the ordinary/inappropriate
33	Affect	flooded
34	Appearance	Issues with Affect
35	Appearance	Sleep/Bed
36	Appearance	Issues with Eye Contact
37	Appearance	Not age appropriate
38	Appearance	Issues with general Health
39	Appearance	Issues with Behavior
40	Appearance	Positive Affect
41	Appearance	Positive Grooming/Hygiene
42	Appearance	Issues with Grooming/Hygiene
43	Appearance	Positive Health
44	Appearance	Issues with Weight
45	Appearance	Normal/no issues
46	Appearance	Hospital attire
47	Appearance	Positive dress
48	Appearance	Issues with dress
49	Appearance	Neutral
50	Appearance	Positive eye contact
51	Appearance	Appropriate Age
52	Appearance	Positive Behavior
53	Association	intact/no issues
54	Association	neutral/unable to categorize
55	Association	circumstantial
56	Association	loose
57	Association	tangential
58	Association	age appropriate
59	Association	improved
60	Association	declined
61	Association	disorganized
62	Association	general issues
63	Attention/ Concentration	intact/no issues
64	Attention/ Concentration	neutral/unable to categorize

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

65	Attention/ Concentration	general issues
66	Attention/ Concentration	improved
67	Attention/ Concentration	declined
68	Attention/ Concentration	age appropriate
69	Attention/ Concentration	issues due to MH
70	Attention/ Concentration	issues due to DD
71	Attention/ Concentration	varying
72	Attitude	positive/appropriate
73	Attitude	neutral/unable to categorize
74	Attitude	age appropriate
75	Attitude	not age appropriate
76	Attitude	negative/inappropriate
77	Attitude	improved
78	Attitude	declined
79	Cognition	normal/no issues
80	Cognition	neutral/unable to categorize
81	Cognition	issues with concentration
82	Cognition	positive concentration
83	Cognition	issues with fund of knowledge
84	Cognition	positive fund of knowledge
85	Cognition	issues with attention
86	Cognition	positive attention
87	Cognition	general issues
88	Executive Functioning	intact/no issues
89	Executive Functioning	neutral/unable to categorize
90	Executive Functioning	some impairment
91	Executive Functioning	much impairment
92	Executive Functioning	improved since last session
93	Executive Functioning	declined since last session
94	Executive Functioning	age appropriate
95	Executive Functioning	below age level
96	Executive Functioning	impaired (due to DD)

For peer review only

Online supplementary material

97	Executive Functioning	impaired (due to TBI)
98	Executive Functioning	impaired (due to other MH issue)
99	Fund of knowledge	intact/no issues
100	Fund of knowledge	neutral/unable to categorize
101	Fund of knowledge	positive vocabulary
102	Fund of knowledge	issues with vocabulary
103	Fund of knowledge	positive history
104	Fund of knowledge	issues with history
105	Fund of knowledge	average
106	Fund of knowledge	above average
107	Fund of knowledge	below average
108	Fund of knowledge	positive current events
109	Fund of knowledge	issues with current events
110	Fund of knowledge	generally limited/issues
111	Fund of knowledge	declined
112	Fund of knowledge	improved
113	Gait and station	normal/no issues
114	Gait and station	neutral/unable to categorize
115	Gait and station	non-ambulatory
116	Gait and station	issues with gait/station
117	Gait and station	assistive device needed
118	Homicidal	not present
119	Homicidal	neutral
120	Homicidal	present
121	Impulse control	good/no issues
122	Impulse control	neutral/unable to categorize
123	Impulse control	limited/some issues
124	Impulse control	poor/serious issues
125	Impulse control	age appropriate
126	Impulse control	not age appropriate
127	Insight	no issues with insight
128	Insight	issues with insight
129	Insight	neutral/unable to categorize
130	Intelligence	normal/no issues with intelligence
131	Intelligence	issues with intelligence
132	Intelligence	neutral/unable to categorize
133	Judgment	no issues with judgement
134	Judgment	issues with judgement
135	Judgment	neutral/unable to categorize
136	Language	intact
137	Language	neutral/unable to categorize

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

138	Language	repetition intact
139	Language	issues with repetition
140	Language	object naming intact
141	Language	issues with object naming
142	Language	impaired
143	Language	non-verbal/mute
144	Language	minimally verbal
145	Language	issues related to DD
146	Level of consciousness	normal; no issues notes
147	Level of consciousness	neutral; unable to categorize
148	Level of consciousness	sleepy/drowsy
149	Level of consciousness	clouded
150	Level of consciousness	issues verbally
151	Level of consciousness	misc. issues
152	Level of consciousness	impaired due to DD/MI
153	Memory	no issues/memory unimpaired
154	Memory	issues/memory impaired
155	Memory	neutral/unable to categorize
156	Mood	Normal; no issues
157	Mood	neutral; unable to categorize
158	Mood	"anxious, tense"
159	Mood	"depressed, sad, despondent"
160	Mood	"irritable, angry"
161	Mood	improved
162	Mood	declined
163	Mood	"elevated, manic"
164	Mood	labile
165	Mood	misc. issues
166	Mood	suicidal
167	Orientation	no orientation issues
168	Orientation	issues with orientation to person
169	Orientation	issues with orientation to place
170	Orientation	issues with orientation to time/date
171	Orientation	issues with orientation to situation
172	Orientation	neutral/unable to categorize
173	Orientation	general orientation issues

For peer review only

Online supplementary material

174	Psychomotor	normal; no issues
175	Psychomotor	neutral; unable to categorize
176	Psychomotor	retarded; slowed
177	Psychomotor	agitated; tense; restless
178	Psychomotor	improved
179	Psychomotor	declined
180	Psychomotor	constricted
181	Psychomotor	tremors
182	Psychomotor	tics
183	Psychomotor	involuntary movements
184	Psychomotor	misc. issues
185	Psychomotor	catatonic
186	Reasoning	normal; no issues
187	Reasoning	neutral; unable to categorize
188	Reasoning	impaired; poor
189	Reasoning	impaired due to DD/cognitive issues
190	Reasoning	impaired due to MI
191	Reasoning	concrete
192	Reasoning	abstract
193	Reasoning	improved
194	Reasoning	misc. issues
195	Sensorium	normal; no issues
196	Sensorium	neutral; unable to categorize
197	Sensorium	clouded; sedate
198	Sensorium	impaired
199	Sensorium	fluctuating
200	Sensorium	impaired due to medications/drugs/alcohol
201	Sensorium	sleepy; lethargic
202	Sleep	"normal, no issues"
203	Sleep	"neutral, unable to categorize"
204	Sleep	maintenance issues (can't stay asleep)
205	Sleep	onset issues (unable to fall asleep)
206	Sleep	nightmares
207	Sleep	insomnia
208	Sleep	hypersomnia
209	Sleep	misc. issues related to sleep
210	Speech	"normal, no issues"
211	Speech	"neutral, unable to categorize"
212	Speech	issues with rate
213	Speech	issues with tone

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

214	Speech	issues with volume
215	Speech	stutter or lisp
216	Speech	misc. issues
217	Speech	impoverished
218	Speech	improved
219	Speech	unintelligible
220	Speech	"mute, non-verbal"
221	Suicidal	"normal, no issues, not present"
222	Suicidal	"neutral, unable to categorize"
223	Suicidal	suicidal ideation
224	Suicidal	suicidal with intent
225	Suicidal	suicidal with plan
226	Suicidal	suicide attempt
227	Suicidal	history of ideation
228	Suicidal	history of attempt
229	Suicidal	self-injurious
230	Suicidal	history of self-injury
231	Suicidal	suicidal ideation with means
232	Violent Thoughts	"normal, no issues"
233	Violent Thoughts	"neutral, unable to categorize"
234	Violent Thoughts	present
235	Violent Thoughts	"present, with intent"
236	Violent Thoughts	present with plan
237	Violent Thoughts	present with means
238	Violent Thoughts	history of violent thoughts
239	Violent Thoughts	history of violent behaviors
240	Violent Thoughts	violent/aggressive behavior

For peer review only

Online supplementary material

eTable 1: Description of U.S. clinics contributing EHR data to NeuroBlu

<Insert at PDF stage>

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

eTable 2: NeuroBlu dataset race distribution

Race	Number of patients	Percentage
White	276,633	49.1%
Unknown	186,517	33.1%
Black or African American	73,504	13.1%
Other Race	13,179	2.3%
Native Hawaiian or Other Pacific Islander	6,712	1.2%
Asian	3,958	0.7%
American Indian or Alaska Native	2,437	0.4%

eTable 3: NeuroBlu dataset age distribution

Age (years)	Number of patients	Percentage
0-9	45,050	8.0%
10-19	116,741	20.7%
20-29	108,863	19.3%
30-39	94,114	16.7%
40-49	81,556	14.5%
50-59	65,498	11.6%
60-69	30,945	5.5%
70-79	15,339	2.7%
80+	4,828	0.9%

eTable 4: NeuroBlu dataset diagnosis distribution

Diagnosis	Number of patients	Percentage
Dementia & Alzheimer's disease	5,029	0.9%
Substance-use disorders	152,790	28.1%
Schizophrenia	14,592	2.7%
Schizoaffective disorder	15,044	2.8%
Bipolar disorder	69,607	12.8%
Major depressive disorder (MDD)	129,120	23.7%
Anxiety disorders	103,923	19.1%
Attention-deficit hyperactivity disorder (ADHD)	53,744	9.9%

For peer review only

Online supplementary material

eTable 5: Gender distribution by diagnosis

Diagnosis	Number of male patients	Number of female patients	Number with unknown gender	Percentage Male	Percentage Female	Percentage Unknown	Total number
Dementia & Alzheimer's disease	2,035	2,962	32	40.5%	58.9%	0.8%	5,029
Substance-use disorders	96,180	56,328	282	62.9%	36.9%	0.0%	152,790
Schizophrenia	9,163	5,401	28	62.8%	37.0%	0.4%	14,592
Schizoaffective disorder	7,807	7,222	15	51.9%	48.0%	0.3%	15,044
Bipolar disorder	27,877	41,642	88	40.0%	59.8%	0.1%	69,607
Major depressive disorder (MDD)	46,603	82,395	122	36.1%	63.8%	0.0%	129,120
Anxiety disorders	38,255	65,569	99	36.8%	63.1%	0.0%	103,923
Attention-deficit hyperactivity disorder (ADHD)	32,781	20,933	30	61.0%	38.9%	0.1%	53,744

For peer review only

Online supplementary material

eTable 6: ICD-9/ICD-10 code definitions for diagnostic category

Diagnosis	ICD code	ICD type	ICD Description
ADHD	314	ICD-9	Attention deficit disorder without mention of hyperactivity
ADHD	314.01	ICD-9	Attention deficit disorder with hyperactivity
ADHD	F90.0	ICD-10	Attention-deficit hyperactivity disorder, predominantly inattentive type
ADHD	F90.1	ICD-10	Attention-deficit hyperactivity disorder, predominantly hyperactive type
ADHD	F90.2	ICD-10	Attention-deficit hyperactivity disorder, combined type
ADHD	F90.8	ICD-10	Attention-deficit hyperactivity disorder, other type
ADHD	F90.9	ICD-10	Attention-deficit hyperactivity disorder, unspecified type
Anxiety disorders	300.01	ICD-9	Panic disorder without agoraphobia
Anxiety disorders	300.02	ICD-9	Generalized anxiety disorder
Anxiety disorders	300.2	ICD-9	Phobia, unspecified
Anxiety disorders	300.21	ICD-9	Agoraphobia with panic disorder
Anxiety disorders	300.22	ICD-9	Agoraphobia without mention of panic attacks
Anxiety disorders	300.23	ICD-9	Social phobia
Anxiety disorders	300.29	ICD-9	Other isolated or specific phobias
Anxiety disorders	300.3	ICD-9	Obsessive-compulsive disorders
Anxiety disorders	309.81	ICD-9	Posttraumatic stress disorder
Anxiety disorders	F40.00	ICD-10	Agoraphobia, unspecified
Anxiety disorders	F40.01	ICD-10	Agoraphobia with panic disorder
Anxiety disorders	F40.02	ICD-10	Agoraphobia without panic disorder
Anxiety disorders	F40.10	ICD-10	Social phobia, unspecified
Anxiety disorders	F40.11	ICD-10	Social phobia, generalized
Anxiety disorders	F40.210	ICD-10	Arachnophobia
Anxiety disorders	F40.218	ICD-10	Other animal type phobia
Anxiety disorders	F40.220	ICD-10	Fear of thunderstorms
Anxiety disorders	F40.228	ICD-10	Other natural environment type phobia

For peer review only

Online supplementary material

Anxiety disorders	F40.230	ICD-10	Fear of blood
Anxiety disorders	F40.231	ICD-10	Fear of injections and transfusions
Anxiety disorders	F40.232	ICD-10	Fear of other medical care
Anxiety disorders	F40.233	ICD-10	Fear of injury
Anxiety disorders	F40.240	ICD-10	Claustrophobia
Anxiety disorders	F40.241	ICD-10	Acrophobia
Anxiety disorders	F40.242	ICD-10	Fear of bridges
Anxiety disorders	F40.243	ICD-10	Fear of flying
Anxiety disorders	F40.248	ICD-10	Other situational type phobia
Anxiety disorders	F40.290	ICD-10	Androphobia
Anxiety disorders	F40.291	ICD-10	Gynephobia
Anxiety disorders	F40.298	ICD-10	Other specified phobia
Anxiety disorders	F40.8	ICD-10	Other phobic anxiety disorders
Anxiety disorders	F40.9	ICD-10	Phobic anxiety disorder, unspecified
Anxiety disorders	F41.0	ICD-10	Panic disorder [episodic paroxysmal anxiety]
Anxiety disorders	F41.1	ICD-10	Generalized anxiety disorder
Anxiety disorders	F42.2	ICD-10	Mixed obsessional thoughts and acts
Anxiety disorders	F42.3	ICD-10	Hoarding disorder
Anxiety disorders	F42.4	ICD-10	Excoriation (skin-picking) disorder
Anxiety disorders	F42.8	ICD-10	Other obsessive-compulsive disorder
Anxiety disorders	F42.9	ICD-10	Obsessive-compulsive disorder, unspecified
Anxiety disorders	F43.10	ICD-10	Post-traumatic stress disorder, unspecified
Anxiety disorders	F43.11	ICD-10	Post-traumatic stress disorder, acute
Anxiety disorders	F43.12	ICD-10	Post-traumatic stress disorder, chronic
Bipolar Disorder	296	ICD-9	Bipolar I disorder, single manic episode, unspecified

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Bipolar Disorder	296.01	ICD-10	Bipolar I disorder, single manic episode, mild
Bipolar Disorder	296.02	ICD-11	Bipolar I disorder, single manic episode, moderate
Bipolar Disorder	296.03	ICD-12	Bipolar I disorder, single manic episode, severe, without mention of psychotic behavior
Bipolar Disorder	296.04	ICD-13	Bipolar I disorder, single manic episode, specified as with psychotic behavior
Bipolar Disorder	296.05	ICD-14	Bipolar I disorder, single manic episode, in partial or unspecified remission
Bipolar Disorder	296.06	ICD-15	Bipolar I disorder, single manic episode, in full remission
Bipolar Disorder	296.4	ICD-9	Bipolar I disorder, most recent episode (or current) manic, unspecified
Bipolar Disorder	296.41	ICD-9	Bipolar I disorder, most recent episode (or current) manic, mild
Bipolar Disorder	296.42	ICD-9	Bipolar I disorder, most recent episode (or current) manic, moderate
Bipolar Disorder	296.43	ICD-9	Bipolar I disorder, most recent episode (or current) manic, severe, without mention of psychotic behavior
Bipolar Disorder	296.44	ICD-9	Bipolar I disorder, most recent episode (or current) manic, severe, specified as with psychotic behavior
Bipolar Disorder	296.45	ICD-9	Bipolar I disorder, most recent episode (or current) manic, in partial or unspecified remission
Bipolar Disorder	296.46	ICD-9	Bipolar I disorder, most recent episode (or current) manic, in full remission
Bipolar Disorder	296.5	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, unspecified
Bipolar Disorder	296.51	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, mild
Bipolar Disorder	296.52	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, moderate
Bipolar Disorder	296.53	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, severe, without mention of psychotic behavior
Bipolar Disorder	296.54	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, severe, specified as with psychotic behavior
Bipolar Disorder	296.55	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, in partial or unspecified remission
Bipolar Disorder	296.56	ICD-9	Bipolar I disorder, most recent episode (or current) depressed, in full remission
Bipolar Disorder	296.6	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, unspecified
Bipolar Disorder	296.61	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, mild
Bipolar Disorder	296.62	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, moderate
Bipolar Disorder	296.63	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, severe, without mention of psychotic behavior
Bipolar Disorder	296.64	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, severe, specified as with psychotic behavior

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Bipolar Disorder	296.65	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, in partial or unspecified remission
Bipolar Disorder	296.66	ICD-9	Bipolar I disorder, most recent episode (or current) mixed, in full remission
Bipolar Disorder	296.7	ICD-9	Bipolar I disorder, most recent episode (or current) unspecified
Bipolar Disorder	296.8	ICD-9	Bipolar disorder, unspecified
Bipolar Disorder	296.89	ICD-9	Other bipolar disorders
Bipolar Disorder	F31.0	ICD-10	Bipolar disorder, current episode hypomanic
Bipolar Disorder	F31.10	ICD-10	Bipolar disorder, current episode manic without psychotic features, unspecified
Bipolar Disorder	F31.11	ICD-10	Bipolar disorder, current episode manic without psychotic features, mild
Bipolar Disorder	F31.12	ICD-10	Bipolar disorder, current episode manic without psychotic features, moderate
Bipolar Disorder	F31.13	ICD-10	Bipolar disorder, current episode manic without psychotic features, severe
Bipolar Disorder	F31.2	ICD-10	Bipolar disorder, current episode manic severe with psychotic features
Bipolar Disorder	F31.30	ICD-10	Bipolar disorder, current episode depressed, mild or moderate severity, unspecified
Bipolar Disorder	F31.31	ICD-10	Bipolar disorder, current episode depressed, mild
Bipolar Disorder	F31.32	ICD-10	Bipolar disorder, current episode depressed, moderate
Bipolar Disorder	F31.4	ICD-10	Bipolar disorder, current episode depressed, severe, without psychotic features
Bipolar Disorder	F31.5	ICD-10	Bipolar disorder, current episode depressed, severe, with psychotic features
Bipolar Disorder	F31.60	ICD-10	Bipolar disorder, current episode mixed, unspecified
Bipolar Disorder	F31.61	ICD-10	Bipolar disorder, current episode mixed, mild
Bipolar Disorder	F31.62	ICD-10	Bipolar disorder, current episode mixed, moderate
Bipolar Disorder	F31.63	ICD-10	Bipolar disorder, current episode mixed, severe, without psychotic features
Bipolar Disorder	F31.64	ICD-10	Bipolar disorder, current episode mixed, severe, with psychotic features
Bipolar Disorder	F31.70	ICD-10	Bipolar disorder, currently in remission, most recent episode unspecified
Bipolar Disorder	F31.71	ICD-10	Bipolar disorder, in partial remission, most recent episode hypomanic
Bipolar Disorder	F31.72	ICD-10	Bipolar disorder, in full remission, most recent episode hypomanic
Bipolar Disorder	F31.73	ICD-10	Bipolar disorder, in partial remission, most recent episode manic

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Bipolar Disorder	F31.74	ICD-10	Bipolar disorder, in full remission, most recent episode manic
Bipolar Disorder	F31.75	ICD-10	Bipolar disorder, in partial remission, most recent episode depressed
Bipolar Disorder	F31.76	ICD-10	Bipolar disorder, in full remission, most recent episode depressed
Bipolar Disorder	F31.77	ICD-10	Bipolar disorder, in partial remission, most recent episode mixed
Bipolar Disorder	F31.78	ICD-10	Bipolar disorder, in full remission, most recent episode mixed
Bipolar Disorder	F31.81	ICD-10	Bipolar II disorder
Bipolar Disorder	F31.89	ICD-10	Other bipolar disorder
Bipolar Disorder	F31.9	ICD-10	Bipolar disorder, unspecified
Dementia and Alzheimer's Disease	290	ICD-9	Senile dementia, uncomplicated
Dementia and Alzheimer's Disease	290.1	ICD-9	Presenile dementia
Dementia and Alzheimer's Disease	290.2	ICD-9	Senile dementia with delusional or depressive features
Dementia and Alzheimer's Disease	290.3	ICD-9	Senile dementia with delirium
Dementia and Alzheimer's Disease	290.4	ICD-9	Vascular dementia
Dementia and Alzheimer's Disease	290.43	ICD-9	Vascular dementia, with depressed mood
Dementia and Alzheimer's Disease	294.1	ICD-9	Dementia in conditions classified elsewhere
Dementia and Alzheimer's Disease	294.8	ICD-9	Other persistent mental disorders due to conditions classified elsewhere
Dementia and Alzheimer's Disease	331	ICD-9	Alzheimer's disease
Dementia and Alzheimer's Disease	F03.90	ICD-10	Unspecified dementia without behavioral disturbance
Dementia and Alzheimer's Disease	F03.91	ICD-10	Unspecified dementia with behavioral disturbance

For peer review only

Online supplementary material

Dementia and Alzheimer's Disease	G30.1	ICD-10	Alzheimer's disease with late onset
Dementia and Alzheimer's Disease	G30.8	ICD-10	Other Alzheimer's disease
MDD	296.2	ICD-9	Major depressive affective disorder, single episode, unspecified
MDD	296.21	ICD-9	Major depressive affective disorder, single episode, mild
MDD	296.22	ICD-9	Major depressive affective disorder, single episode, moderate
MDD	296.23	ICD-9	Major depressive affective disorder, single episode, severe, without mention of psychotic behavior
MDD	296.24	ICD-9	Major depressive affective disorder, single episode, severe, specified as with psychotic behavior
MDD	296.25	ICD-9	Major depressive affective disorder, single episode, in partial or unspecified remission
MDD	296.26	ICD-9	Major depressive affective disorder, single episode, in full remission
MDD	296.3	ICD-9	Major depressive affective disorder, recurrent episode, unspecified
MDD	296.31	ICD-9	Major depressive affective disorder, recurrent episode, mild
MDD	296.32	ICD-9	Major depressive affective disorder, recurrent episode, moderate
MDD	296.33	ICD-9	Major depressive affective disorder, recurrent episode, severe, without mention of psychotic behavior
MDD	296.34	ICD-9	Major depressive affective disorder, recurrent episode, severe, specified as with psychotic behavior
MDD	296.35	ICD-9	Major depressive affective disorder, recurrent episode, in partial or unspecified remission
MDD	296.36	ICD-9	Major depressive affective disorder, recurrent episode, in full remission
MDD	F32.0	ICD-10	Major depressive disorder, single episode, mild
MDD	F32.1	ICD-10	Major depressive disorder, single episode, moderate
MDD	F32.2	ICD-10	Major depressive disorder, single episode, severe without psychotic features
MDD	F32.3	ICD-10	Major depressive disorder, single episode, severe with psychotic features
MDD	F32.4	ICD-10	Major depressive disorder, single episode, in partial remission
MDD	F32.5	ICD-10	Major depressive disorder, single episode, in full remission
MDD	F32.81	ICD-10	Premenstrual dysphoric disorder
MDD	F32.89	ICD-10	Other specified depressive episodes
MDD	F32.9	ICD-10	Major depressive disorder, single episode, unspecified
MDD	F33.0	ICD-10	Major depressive disorder, recurrent, mild
MDD	F33.1	ICD-10	Major depressive disorder, recurrent, moderate

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

MDD	F33.2	ICD-10	Major depressive disorder, recurrent severe without psychotic features
MDD	F33.3	ICD-10	Major depressive disorder, recurrent, severe with psychotic symptoms
MDD	F33.40	ICD-10	Major depressive disorder, recurrent, in remission, unspecified
MDD	F33.41	ICD-10	Major depressive disorder, recurrent, in partial remission
MDD	F33.42	ICD-10	Major depressive disorder, recurrent, in full remission
MDD	F33.8	ICD-10	Other recurrent depressive disorders
MDD	F33.9	ICD-10	Major depressive disorder, recurrent, unspecified
Schizoaffective disorder	295.7	ICD-9	Schizoaffective disorder, unspecified
Schizoaffective disorder	295.71	ICD-9	Schizoaffective disorder, subchronic
Schizoaffective disorder	295.72	ICD-9	Schizoaffective disorder, chronic
Schizoaffective disorder	295.73	ICD-9	Schizoaffective disorder, subchronic with acute exacerbation
Schizoaffective disorder	295.74	ICD-9	Schizoaffective disorder, chronic with acute exacerbation
Schizoaffective disorder	295.75	ICD-9	Schizoaffective disorder, in remission
Schizoaffective disorder	F25.0	ICD-10	Schizoaffective disorder, bipolar type
Schizoaffective disorder	F25.1	ICD-10	Schizoaffective disorder, depressive type
Schizoaffective disorder	F25.8	ICD-10	Other schizoaffective disorders
Schizoaffective disorder	F25.9	ICD-10	Schizoaffective disorder, unspecified
Schizophrenia	295	ICD-9	Simple type schizophrenia, unspecified
Schizophrenia	295.01	ICD-9	Simple type schizophrenia, subchronic
Schizophrenia	295.02	ICD-9	Simple type schizophrenia, chronic
Schizophrenia	295.03	ICD-9	Simple type schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.04	ICD-9	Simple type schizophrenia, chronic with acute exacerbation
Schizophrenia	295.05	ICD-9	Simple type schizophrenia, in remission
Schizophrenia	295.1	ICD-9	Disorganized type schizophrenia, unspecified
Schizophrenia	295.11	ICD-9	Disorganized type schizophrenia, subchronic
Schizophrenia	295.12	ICD-9	Disorganized type schizophrenia, chronic
Schizophrenia	295.13	ICD-9	Disorganized type schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.14	ICD-9	Disorganized type schizophrenia, chronic with acute exacerbation
Schizophrenia	295.15	ICD-9	Disorganized type schizophrenia, in remission
Schizophrenia	295.2	ICD-9	Catatonic type schizophrenia, unspecified

For peer review only

Online supplementary material

Schizophrenia	295.21	ICD-9	Catatonic type schizophrenia, subchronic
Schizophrenia	295.22	ICD-9	Catatonic type schizophrenia, chronic
Schizophrenia	295.23	ICD-9	Catatonic type schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.24	ICD-9	Catatonic type schizophrenia, chronic with acute exacerbation
Schizophrenia	295.25	ICD-9	Catatonic type schizophrenia, in remission
Schizophrenia	295.3	ICD-9	Paranoid type schizophrenia, unspecified
Schizophrenia	295.31	ICD-9	Paranoid type schizophrenia, subchronic
Schizophrenia	295.32	ICD-9	Paranoid type schizophrenia, chronic
Schizophrenia	295.33	ICD-9	Paranoid type schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.34	ICD-9	Paranoid type schizophrenia, chronic with acute exacerbation
Schizophrenia	295.35	ICD-9	Paranoid type schizophrenia, in remission
Schizophrenia	295.4	ICD-9	Schizophreniform disorder, unspecified
Schizophrenia	295.41	ICD-9	Schizophreniform disorder, subchronic
Schizophrenia	295.42	ICD-9	Schizophreniform disorder, chronic
Schizophrenia	295.43	ICD-9	Schizophreniform disorder, subchronic with acute exacerbation
Schizophrenia	295.44	ICD-9	Schizophreniform disorder, chronic with acute exacerbation
Schizophrenia	295.45	ICD-9	Schizophreniform disorder, in remission
Schizophrenia	295.5	ICD-9	Latent schizophrenia, unspecified
Schizophrenia	295.51	ICD-9	Latent schizophrenia, subchronic
Schizophrenia	295.52	ICD-9	Latent schizophrenia, chronic
Schizophrenia	295.53	ICD-9	Latent schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.54	ICD-9	Latent schizophrenia, chronic with acute exacerbation
Schizophrenia	295.55	ICD-9	Latent schizophrenia, in remission
Schizophrenia	295.6	ICD-9	Schizophrenic disorders, residual type, unspecified
Schizophrenia	295.61	ICD-9	Schizophrenic disorders, residual type, subchronic
Schizophrenia	295.62	ICD-9	Schizophrenic disorders, residual type, chronic
Schizophrenia	295.63	ICD-9	Schizophrenic disorders, residual type, subchronic with acute exacerbation
Schizophrenia	295.64	ICD-9	Schizophrenic disorders, residual type, chronic with acute exacerbation
Schizophrenia	295.65	ICD-9	Schizophrenic disorders, residual type, in remission
Schizophrenia	295.8	ICD-9	Other specified types of schizophrenia, unspecified
Schizophrenia	295.81	ICD-9	Other specified types of schizophrenia, subchronic
Schizophrenia	295.82	ICD-9	Other specified types of schizophrenia, chronic
Schizophrenia	295.83	ICD-9	Other specified types of schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.84	ICD-9	Other specified types of schizophrenia, chronic with acute exacerbation
Schizophrenia	295.85	ICD-9	Other specified types of schizophrenia, in remission

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Schizophrenia	295.9	ICD-9	Unspecified schizophrenia, unspecified
Schizophrenia	295.91	ICD-9	Unspecified schizophrenia, subchronic
Schizophrenia	295.92	ICD-9	Unspecified schizophrenia, chronic
Schizophrenia	295.93	ICD-9	Unspecified schizophrenia, subchronic with acute exacerbation
Schizophrenia	295.94	ICD-9	Unspecified schizophrenia, chronic with acute exacerbation
Schizophrenia	295.95	ICD-9	Unspecified schizophrenia, in remission
Schizophrenia	F20.0	ICD-10	Paranoid schizophrenia
Schizophrenia	F20.1	ICD-10	Disorganized schizophrenia
Schizophrenia	F20.2	ICD-10	Catatonic schizophrenia
Schizophrenia	F20.3	ICD-10	Undifferentiated schizophrenia
Schizophrenia	F20.5	ICD-10	Residual schizophrenia
Schizophrenia	F20.81	ICD-10	Schizophreniform disorder
Schizophrenia	F20.89	ICD-10	Other schizophrenia
Schizophrenia	F20.9	ICD-10	Schizophrenia, unspecified
Substance-use -disorders	291	ICD-9	Alcohol withdrawal delirium
Substance-use -disorders	291.1	ICD-9	Alcohol-induced persisting amnestic disorder
Substance-use -disorders	291.2	ICD-9	Alcohol-induced persisting dementia
Substance-use -disorders	291.3	ICD-9	Alcohol-induced psychotic disorder with hallucinations
Substance-use -disorders	291.4	ICD-9	Idiosyncratic alcohol intoxication
Substance-use -disorders	291.5	ICD-9	Alcohol-induced psychotic disorder with delusions
Substance-use -disorders	291.81	ICD-9	Alcohol withdrawal
Substance-use -disorders	291.82	ICD-9	Alcohol induced sleep disorders
Substance-use -disorders	291.89	ICD-9	Other alcohol-induced mental disorders
Substance-use -disorders	291.9	ICD-9	Unspecified alcohol-induced mental disorders
Substance-use -disorders	292	ICD-9	Drug withdrawal
Substance-use -disorders	292.11	ICD-9	Drug-induced psychotic disorder with delusions
Substance-use -disorders	292.12	ICD-9	Drug-induced psychotic disorder with hallucinations
Substance-use -disorders	292.2	ICD-9	Pathological drug intoxication
Substance-use -disorders	292.81	ICD-9	Drug-induced delirium
Substance-use -disorders	292.82	ICD-9	Drug-induced persisting dementia

For peer review only

Online supplementary material

Substance-use -disorders	292.83	ICD-9	Drug-induced persisting amnestic disorder
Substance-use -disorders	292.84	ICD-9	Drug-induced mood disorder
Substance-use -disorders	292.85	ICD-9	Drug induced sleep disorders
Substance-use -disorders	292.89	ICD-9	Other specified drug-induced mental disorders
Substance-use -disorders	292.9	ICD-9	Unspecified drug-induced mental disorder
Substance-use -disorders	303	ICD-9	Acute alcoholic intoxication in alcoholism, unspecified
Substance-use -disorders	303.01	ICD-9	Acute alcoholic intoxication in alcoholism, continuous
Substance-use -disorders	303.02	ICD-9	Acute alcoholic intoxication in alcoholism, episodic
Substance-use -disorders	303.03	ICD-9	Acute alcoholic intoxication in alcoholism, in remission
Substance-use -disorders	303.9	ICD-9	Other and unspecified alcohol dependence, unspecified
Substance-use -disorders	303.91	ICD-9	Other and unspecified alcohol dependence, continuous
Substance-use -disorders	303.92	ICD-9	Other and unspecified alcohol dependence, episodic
Substance-use -disorders	303.93	ICD-9	Other and unspecified alcohol dependence, in remission
Substance-use -disorders	304	ICD-9	Opioid type dependence, unspecified
Substance-use -disorders	304.01	ICD-9	Opioid type dependence, continuous
Substance-use -disorders	304.02	ICD-9	Opioid type dependence, episodic
Substance-use -disorders	304.03	ICD-9	Opioid type dependence, in remission
Substance-use -disorders	304.1	ICD-9	Sedative, hypnotic or anxiolytic dependence, unspecified
Substance-use -disorders	304.11	ICD-9	Sedative, hypnotic or anxiolytic dependence, continuous
Substance-use -disorders	304.12	ICD-9	Sedative, hypnotic or anxiolytic dependence, episodic
Substance-use -disorders	304.13	ICD-9	Sedative, hypnotic or anxiolytic dependence, in remission
Substance-use -disorders	304.2	ICD-9	Cocaine dependence, unspecified
Substance-use -disorders	304.21	ICD-9	Cocaine dependence, continuous
Substance-use -disorders	304.22	ICD-9	Cocaine dependence, episodic
Substance-use -disorders	304.23	ICD-9	Cocaine dependence, in remission

For peer review only

Online supplementary material

Substance-use -disorders	304.3	ICD-9	Cannabis dependence, unspecified
Substance-use -disorders	304.31	ICD-9	Cannabis dependence, continuous
Substance-use -disorders	304.32	ICD-9	Cannabis dependence, episodic
Substance-use -disorders	304.33	ICD-9	Cannabis dependence, in remission
Substance-use -disorders	304.4	ICD-9	Amphetamine and other psychostimulant dependence, unspecified
Substance-use -disorders	304.41	ICD-9	Amphetamine and other psychostimulant dependence, continuous
Substance-use -disorders	304.42	ICD-9	Amphetamine and other psychostimulant dependence, episodic
Substance-use -disorders	304.43	ICD-9	Amphetamine and other psychostimulant dependence, in remission
Substance-use -disorders	304.5	ICD-9	Hallucinogen dependence, unspecified
Substance-use -disorders	304.51	ICD-9	Hallucinogen dependence, continuous
Substance-use -disorders	304.52	ICD-9	Hallucinogen dependence, episodic
Substance-use -disorders	304.53	ICD-9	Hallucinogen dependence, in remission
Substance-use -disorders	304.6	ICD-9	Other specified drug dependence, unspecified
Substance-use -disorders	304.61	ICD-9	Other specified drug dependence, continuous
Substance-use -disorders	304.62	ICD-9	Other specified drug dependence, episodic
Substance-use -disorders	304.63	ICD-9	Other specified drug dependence, in remission
Substance-use -disorders	304.7	ICD-9	Combinations of opioid type drug with any other drug dependence, unspecified
Substance-use -disorders	304.71	ICD-9	Combinations of opioid type drug with any other drug dependence, continuous
Substance-use -disorders	304.72	ICD-9	Combinations of opioid type drug with any other drug dependence, episodic
Substance-use -disorders	304.73	ICD-9	Combinations of opioid type drug with any other drug dependence, in remission
Substance-use -disorders	304.8	ICD-9	Combinations of drug dependence excluding opioid type drug, unspecified
Substance-use -disorders	304.81	ICD-9	Combinations of drug dependence excluding opioid type drug, continuous
Substance-use -disorders	304.82	ICD-9	Combinations of drug dependence excluding opioid type drug, episodic
Substance-use -disorders	304.83	ICD-9	Combinations of drug dependence excluding opioid type drug, in remission
Substance-use -disorders	304.9	ICD-9	Unspecified drug dependence, unspecified

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	304.91	ICD-9	Unspecified drug dependence, continuous
Substance-use -disorders	304.92	ICD-9	Unspecified drug dependence, episodic
Substance-use -disorders	304.93	ICD-9	Unspecified drug dependence, in remission
Substance-use -disorders	305	ICD-9	Alcohol abuse, unspecified
Substance-use -disorders	305.01	ICD-9	Alcohol abuse, continuous
Substance-use -disorders	305.02	ICD-9	Alcohol abuse, episodic
Substance-use -disorders	305.03	ICD-9	Alcohol abuse, in remission
Substance-use -disorders	305.1	ICD-9	Tobacco use disorder
Substance-use -disorders	305.2	ICD-9	Cannabis abuse, unspecified
Substance-use -disorders	305.21	ICD-9	Cannabis abuse, continuous
Substance-use -disorders	305.22	ICD-9	Cannabis abuse, episodic
Substance-use -disorders	305.23	ICD-9	Cannabis abuse, in remission
Substance-use -disorders	305.3	ICD-9	Hallucinogen abuse, unspecified
Substance-use -disorders	305.31	ICD-9	Hallucinogen abuse, continuous
Substance-use -disorders	305.32	ICD-9	Hallucinogen abuse, episodic
Substance-use -disorders	305.33	ICD-9	Hallucinogen abuse, in remission
Substance-use -disorders	305.4	ICD-9	Sedative, hypnotic or anxiolytic abuse, unspecified
Substance-use -disorders	305.41	ICD-9	Sedative, hypnotic or anxiolytic abuse, continuous
Substance-use -disorders	305.42	ICD-9	Sedative, hypnotic or anxiolytic abuse, episodic
Substance-use -disorders	305.43	ICD-9	Sedative, hypnotic or anxiolytic abuse, in remission
Substance-use -disorders	305.5	ICD-9	Opioid abuse, unspecified
Substance-use -disorders	305.51	ICD-9	Opioid abuse, continuous
Substance-use -disorders	305.52	ICD-9	Opioid abuse, episodic
Substance-use -disorders	305.53	ICD-9	Opioid abuse, in remission
Substance-use -disorders	305.6	ICD-9	Cocaine abuse, unspecified

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	305.61	ICD-9	Cocaine abuse, continuous
Substance-use -disorders	305.62	ICD-9	Cocaine abuse, episodic
Substance-use -disorders	305.63	ICD-9	Cocaine abuse, in remission
Substance-use -disorders	305.7	ICD-9	Amphetamine or related acting sympathomimetic abuse, unspecified
Substance-use -disorders	305.71	ICD-9	Amphetamine or related acting sympathomimetic abuse, continuous
Substance-use -disorders	305.72	ICD-9	Amphetamine or related acting sympathomimetic abuse, episodic
Substance-use -disorders	305.73	ICD-9	Amphetamine or related acting sympathomimetic abuse, in remission
Substance-use -disorders	305.8	ICD-9	Antidepressant type abuse, unspecified
Substance-use -disorders	305.81	ICD-9	Antidepressant type abuse, continuous
Substance-use -disorders	305.82	ICD-9	Antidepressant type abuse, episodic
Substance-use -disorders	305.83	ICD-9	Antidepressant type abuse, in remission
Substance-use -disorders	305.9	ICD-9	Other, mixed, or unspecified drug abuse, unspecified
Substance-use -disorders	305.91	ICD-9	Other, mixed, or unspecified drug abuse, continuous
Substance-use -disorders	305.92	ICD-9	Other, mixed, or unspecified drug abuse, episodic
Substance-use -disorders	305.93	ICD-9	Other, mixed, or unspecified drug abuse, in remission
Substance-use -disorders	F10.10	ICD-10	Alcohol abuse, uncomplicated
Substance-use -disorders	F10.11	ICD-10	Alcohol abuse, in remission
Substance-use -disorders	F10.120	ICD-10	Alcohol abuse with intoxication, uncomplicated
Substance-use -disorders	F10.121	ICD-10	Alcohol abuse with intoxication delirium
Substance-use -disorders	F10.129	ICD-10	Alcohol abuse with intoxication, unspecified
Substance-use -disorders	F10.14	ICD-10	Alcohol abuse with alcohol-induced mood disorder
Substance-use -disorders	F10.150	ICD-10	Alcohol abuse with alcohol-induced psychotic disorder with delusions
Substance-use -disorders	F10.151	ICD-10	Alcohol abuse with alcohol-induced psychotic disorder with hallucinations
Substance-use -disorders	F10.159	ICD-10	Alcohol abuse with alcohol-induced psychotic disorder, unspecified
Substance-use -disorders	F10.180	ICD-10	Alcohol abuse with alcohol-induced anxiety disorder

For peer review only

Online supplementary material

Substance-use -disorders	F10.181	ICD-10	Alcohol abuse with alcohol-induced sexual dysfunction
Substance-use -disorders	F10.182	ICD-10	Alcohol abuse with alcohol-induced sleep disorder
Substance-use -disorders	F10.188	ICD-10	Alcohol abuse with other alcohol-induced disorder
Substance-use -disorders	F10.19	ICD-10	Alcohol abuse with unspecified alcohol-induced disorder
Substance-use -disorders	F10.20	ICD-10	Alcohol dependence, uncomplicated
Substance-use -disorders	F10.21	ICD-10	Alcohol dependence, in remission
Substance-use -disorders	F10.220	ICD-10	Alcohol dependence with intoxication, uncomplicated
Substance-use -disorders	F10.221	ICD-10	Alcohol dependence with intoxication delirium
Substance-use -disorders	F10.229	ICD-10	Alcohol dependence with intoxication, unspecified
Substance-use -disorders	F10.230	ICD-10	Alcohol dependence with withdrawal, uncomplicated
Substance-use -disorders	F10.231	ICD-10	Alcohol dependence with withdrawal delirium
Substance-use -disorders	F10.232	ICD-10	Alcohol dependence with withdrawal with perceptual disturbance
Substance-use -disorders	F10.239	ICD-10	Alcohol dependence with withdrawal, unspecified
Substance-use -disorders	F10.24	ICD-10	Alcohol dependence with alcohol-induced mood disorder
Substance-use -disorders	F10.250	ICD-10	Alcohol dependence with alcohol-induced psychotic disorder with delusions
Substance-use -disorders	F10.251	ICD-10	Alcohol dependence with alcohol-induced psychotic disorder with hallucinations
Substance-use -disorders	F10.259	ICD-10	Alcohol dependence with alcohol-induced psychotic disorder, unspecified
Substance-use -disorders	F10.26	ICD-10	Alcohol dependence with alcohol-induced persisting amnestic disorder
Substance-use -disorders	F10.27	ICD-10	Alcohol dependence with alcohol-induced persisting dementia
Substance-use -disorders	F10.280	ICD-10	Alcohol dependence with alcohol-induced anxiety disorder
Substance-use -disorders	F10.281	ICD-10	Alcohol dependence with alcohol-induced sexual dysfunction
Substance-use -disorders	F10.282	ICD-10	Alcohol dependence with alcohol-induced sleep disorder
Substance-use -disorders	F10.288	ICD-10	Alcohol dependence with other alcohol-induced disorder
Substance-use -disorders	F10.29	ICD-10	Alcohol dependence with unspecified alcohol-induced disorder
Substance-use -disorders	F10.920	ICD-10	Alcohol use, unspecified with intoxication, uncomplicated

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F10.921	ICD-10	Alcohol use, unspecified with intoxication delirium
Substance-use -disorders	F10.929	ICD-10	Alcohol use, unspecified with intoxication, unspecified
Substance-use -disorders	F10.94	ICD-10	Alcohol use, unspecified with alcohol-induced mood disorder
Substance-use -disorders	F10.950	ICD-10	Alcohol use, unspecified with alcohol-induced psychotic disorder with delusions
Substance-use -disorders	F10.951	ICD-10	Alcohol use, unspecified with alcohol-induced psychotic disorder with hallucinations
Substance-use -disorders	F10.959	ICD-10	Alcohol use, unspecified with alcohol-induced psychotic disorder, unspecified
Substance-use -disorders	F10.96	ICD-10	Alcohol use, unspecified with alcohol-induced persisting amnestic disorder
Substance-use -disorders	F10.97	ICD-10	Alcohol use, unspecified with alcohol-induced persisting dementia
Substance-use -disorders	F10.980	ICD-10	Alcohol use, unspecified with alcohol-induced anxiety disorder
Substance-use -disorders	F10.981	ICD-10	Alcohol use, unspecified with alcohol-induced sexual dysfunction
Substance-use -disorders	F10.982	ICD-10	Alcohol use, unspecified with alcohol-induced sleep disorder
Substance-use -disorders	F10.988	ICD-10	Alcohol use, unspecified with other alcohol-induced disorder
Substance-use -disorders	F10.99	ICD-10	Alcohol use, unspecified with unspecified alcohol-induced disorder
Substance-use -disorders	F11.10	ICD-10	Opioid abuse, uncomplicated
Substance-use -disorders	F11.11	ICD-10	Opioid abuse, in remission
Substance-use -disorders	F11.120	ICD-10	Opioid abuse with intoxication, uncomplicated
Substance-use -disorders	F11.121	ICD-10	Opioid abuse with intoxication delirium
Substance-use -disorders	F11.122	ICD-10	Opioid abuse with intoxication with perceptual disturbance
Substance-use -disorders	F11.129	ICD-10	Opioid abuse with intoxication, unspecified
Substance-use -disorders	F11.14	ICD-10	Opioid abuse with opioid-induced mood disorder
Substance-use -disorders	F11.150	ICD-10	Opioid abuse with opioid-induced psychotic disorder with delusions
Substance-use -disorders	F11.151	ICD-10	Opioid abuse with opioid-induced psychotic disorder with hallucinations
Substance-use -disorders	F11.159	ICD-10	Opioid abuse with opioid-induced psychotic disorder, unspecified
Substance-use -disorders	F11.181	ICD-10	Opioid abuse with opioid-induced sexual dysfunction
Substance-use -disorders	F11.182	ICD-10	Opioid abuse with opioid-induced sleep disorder

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F11.188	ICD-10	Opioid abuse with other opioid-induced disorder
Substance-use -disorders	F11.19	ICD-10	Opioid abuse with unspecified opioid-induced disorder
Substance-use -disorders	F11.20	ICD-10	Opioid dependence, uncomplicated
Substance-use -disorders	F11.21	ICD-10	Opioid dependence, in remission
Substance-use -disorders	F11.220	ICD-10	Opioid dependence with intoxication, uncomplicated
Substance-use -disorders	F11.221	ICD-10	Opioid dependence with intoxication delirium
Substance-use -disorders	F11.222	ICD-10	Opioid dependence with intoxication with perceptual disturbance
Substance-use -disorders	F11.229	ICD-10	Opioid dependence with intoxication, unspecified
Substance-use -disorders	F11.23	ICD-10	Opioid dependence with withdrawal
Substance-use -disorders	F11.24	ICD-10	Opioid dependence with opioid-induced mood disorder
Substance-use -disorders	F11.250	ICD-10	Opioid dependence with opioid-induced psychotic disorder with delusions
Substance-use -disorders	F11.251	ICD-10	Opioid dependence with opioid-induced psychotic disorder with hallucinations
Substance-use -disorders	F11.259	ICD-10	Opioid dependence with opioid-induced psychotic disorder, unspecified
Substance-use -disorders	F11.281	ICD-10	Opioid dependence with opioid-induced sexual dysfunction
Substance-use -disorders	F11.282	ICD-10	Opioid dependence with opioid-induced sleep disorder
Substance-use -disorders	F11.288	ICD-10	Opioid dependence with other opioid-induced disorder
Substance-use -disorders	F11.29	ICD-10	Opioid dependence with unspecified opioid-induced disorder
Substance-use -disorders	F11.90	ICD-10	Opioid use, unspecified, uncomplicated
Substance-use -disorders	F11.920	ICD-10	Opioid use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F11.921	ICD-10	Opioid use, unspecified with intoxication delirium
Substance-use -disorders	F11.922	ICD-10	Opioid use, unspecified with intoxication with perceptual disturbance
Substance-use -disorders	F11.929	ICD-10	Opioid use, unspecified with intoxication, unspecified
Substance-use -disorders	F11.93	ICD-10	Opioid use, unspecified with withdrawal
Substance-use -disorders	F11.94	ICD-10	Opioid use, unspecified with opioid-induced mood disorder
Substance-use -disorders	F11.950	ICD-10	Opioid use, unspecified with opioid-induced psychotic disorder with delusions

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F11.951	ICD-10	Opioid use, unspecified with opioid-induced psychotic disorder with hallucinations
Substance-use -disorders	F11.959	ICD-10	Opioid use, unspecified with opioid-induced psychotic disorder, unspecified
Substance-use -disorders	F11.981	ICD-10	Opioid use, unspecified with opioid-induced sexual dysfunction
Substance-use -disorders	F11.982	ICD-10	Opioid use, unspecified with opioid-induced sleep disorder
Substance-use -disorders	F11.988	ICD-10	Opioid use, unspecified with other opioid-induced disorder
Substance-use -disorders	F11.99	ICD-10	Opioid use, unspecified with unspecified opioid-induced disorder
Substance-use -disorders	F12.10	ICD-10	Cannabis abuse, uncomplicated
Substance-use -disorders	F12.11	ICD-10	Cannabis abuse, in remission
Substance-use -disorders	F12.120	ICD-10	Cannabis abuse with intoxication, uncomplicated
Substance-use -disorders	F12.121	ICD-10	Cannabis abuse with intoxication delirium
Substance-use -disorders	F12.122	ICD-10	Cannabis abuse with intoxication with perceptual disturbance
Substance-use -disorders	F12.129	ICD-10	Cannabis abuse with intoxication, unspecified
Substance-use -disorders	F12.150	ICD-10	Cannabis abuse with psychotic disorder with delusions
Substance-use -disorders	F12.151	ICD-10	Cannabis abuse with psychotic disorder with hallucinations
Substance-use -disorders	F12.159	ICD-10	Cannabis abuse with psychotic disorder, unspecified
Substance-use -disorders	F12.180	ICD-10	Cannabis abuse with cannabis-induced anxiety disorder
Substance-use -disorders	F12.188	ICD-10	Cannabis abuse with other cannabis-induced disorder
Substance-use -disorders	F12.19	ICD-10	Cannabis abuse with unspecified cannabis-induced disorder
Substance-use -disorders	F12.20	ICD-10	Cannabis dependence, uncomplicated
Substance-use -disorders	F12.21	ICD-10	Cannabis dependence, in remission
Substance-use -disorders	F12.220	ICD-10	Cannabis dependence with intoxication, uncomplicated
Substance-use -disorders	F12.221	ICD-10	Cannabis dependence with intoxication delirium
Substance-use -disorders	F12.222	ICD-10	Cannabis dependence with intoxication with perceptual disturbance
Substance-use -disorders	F12.229	ICD-10	Cannabis dependence with intoxication, unspecified
Substance-use -disorders	F12.250	ICD-10	Cannabis dependence with psychotic disorder with delusions

For peer review only

Online supplementary material

Substance-use -disorders	F12.251	ICD-10	Cannabis dependence with psychotic disorder with hallucinations
Substance-use -disorders	F12.259	ICD-10	Cannabis dependence with psychotic disorder, unspecified
Substance-use -disorders	F12.280	ICD-10	Cannabis dependence with cannabis-induced anxiety disorder
Substance-use -disorders	F12.288	ICD-10	Cannabis dependence with other cannabis-induced disorder
Substance-use -disorders	F12.29	ICD-10	Cannabis dependence with unspecified cannabis-induced disorder
Substance-use -disorders	F12.90	ICD-10	Cannabis use, unspecified, uncomplicated
Substance-use -disorders	F12.920	ICD-10	Cannabis use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F12.921	ICD-10	Cannabis use, unspecified with intoxication delirium
Substance-use -disorders	F12.922	ICD-10	Cannabis use, unspecified with intoxication with perceptual disturbance
Substance-use -disorders	F12.929	ICD-10	Cannabis use, unspecified with intoxication, unspecified
Substance-use -disorders	F12.950	ICD-10	Cannabis use, unspecified with psychotic disorder with delusions
Substance-use -disorders	F12.951	ICD-10	Cannabis use, unspecified with psychotic disorder with hallucinations
Substance-use -disorders	F12.959	ICD-10	Cannabis use, unspecified with psychotic disorder, unspecified
Substance-use -disorders	F12.980	ICD-10	Cannabis use, unspecified with anxiety disorder
Substance-use -disorders	F12.988	ICD-10	Cannabis use, unspecified with other cannabis-induced disorder
Substance-use -disorders	F12.99	ICD-10	Cannabis use, unspecified with unspecified cannabis-induced disorder
Substance-use -disorders	F13.10	ICD-10	Sedative, hypnotic or anxiolytic abuse, uncomplicated
Substance-use -disorders	F13.11	ICD-10	Sedative, hypnotic or anxiolytic abuse, in remission
Substance-use -disorders	F13.120	ICD-10	Sedative, hypnotic or anxiolytic abuse with intoxication, uncomplicated
Substance-use -disorders	F13.121	ICD-10	Sedative, hypnotic or anxiolytic abuse with intoxication delirium
Substance-use -disorders	F13.129	ICD-10	Sedative, hypnotic or anxiolytic abuse with intoxication, unspecified
Substance-use -disorders	F13.14	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced mood disorder
Substance-use -disorders	F13.150	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced psychotic disorder with delusions
Substance-use -disorders	F13.151	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced psychotic disorder with hallucinations

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F13.159	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced psychotic disorder, unspecified
Substance-use -disorders	F13.180	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced anxiety disorder
Substance-use -disorders	F13.181	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced sexual dysfunction
Substance-use -disorders	F13.182	ICD-10	Sedative, hypnotic or anxiolytic abuse with sedative, hypnotic or anxiolytic-induced sleep disorder
Substance-use -disorders	F13.188	ICD-10	Sedative, hypnotic or anxiolytic abuse with other sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F13.19	ICD-10	Sedative, hypnotic or anxiolytic abuse with unspecified sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F13.20	ICD-10	Sedative, hypnotic or anxiolytic dependence, uncomplicated
Substance-use -disorders	F13.21	ICD-10	Sedative, hypnotic or anxiolytic dependence, in remission
Substance-use -disorders	F13.220	ICD-10	Sedative, hypnotic or anxiolytic dependence with intoxication, uncomplicated
Substance-use -disorders	F13.221	ICD-10	Sedative, hypnotic or anxiolytic dependence with intoxication delirium
Substance-use -disorders	F13.229	ICD-10	Sedative, hypnotic or anxiolytic dependence with intoxication, unspecified
Substance-use -disorders	F13.230	ICD-10	Sedative, hypnotic or anxiolytic dependence with withdrawal, uncomplicated
Substance-use -disorders	F13.231	ICD-10	Sedative, hypnotic or anxiolytic dependence with withdrawal delirium
Substance-use -disorders	F13.232	ICD-10	Sedative, hypnotic or anxiolytic dependence with withdrawal with perceptual disturbance
Substance-use -disorders	F13.239	ICD-10	Sedative, hypnotic or anxiolytic dependence with withdrawal, unspecified
Substance-use -disorders	F13.24	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced mood disorder
Substance-use -disorders	F13.250	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced psychotic disorder with delusions
Substance-use -disorders	F13.251	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced psychotic disorder with hallucinations
Substance-use -disorders	F13.259	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced psychotic disorder, unspecified
Substance-use -disorders	F13.26	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced persisting amnestic disorder
Substance-use -disorders	F13.27	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced persisting dementia
Substance-use -disorders	F13.280	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced anxiety disorder

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F13.281	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced sexual dysfunction
Substance-use -disorders	F13.282	ICD-10	Sedative, hypnotic or anxiolytic dependence with sedative, hypnotic or anxiolytic-induced sleep disorder
Substance-use -disorders	F13.288	ICD-10	Sedative, hypnotic or anxiolytic dependence with other sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F13.29	ICD-10	Sedative, hypnotic or anxiolytic dependence with unspecified sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F13.90	ICD-10	Sedative, hypnotic, or anxiolytic use, unspecified, uncomplicated
Substance-use -disorders	F13.920	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F13.921	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with intoxication delirium
Substance-use -disorders	F13.929	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with intoxication, unspecified
Substance-use -disorders	F13.930	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with withdrawal, uncomplicated
Substance-use -disorders	F13.931	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with withdrawal delirium
Substance-use -disorders	F13.932	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with withdrawal with perceptual disturbances
Substance-use -disorders	F13.939	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with withdrawal, unspecified
Substance-use -disorders	F13.94	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced mood disorder
Substance-use -disorders	F13.950	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced psychotic disorder with delusions
Substance-use -disorders	F13.951	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced psychotic disorder with hallucinations
Substance-use -disorders	F13.959	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced psychotic disorder, unspecified
Substance-use -disorders	F13.96	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced persisting amnestic disorder
Substance-use -disorders	F13.97	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced persisting dementia
Substance-use -disorders	F13.980	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced anxiety disorder
Substance-use -disorders	F13.981	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced sexual dysfunction
Substance-use -disorders	F13.982	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with sedative, hypnotic or anxiolytic-induced sleep disorder

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F13.988	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with other sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F13.99	ICD-10	Sedative, hypnotic or anxiolytic use, unspecified with unspecified sedative, hypnotic or anxiolytic-induced disorder
Substance-use -disorders	F14.10	ICD-10	Cocaine abuse, uncomplicated
Substance-use -disorders	F14.11	ICD-10	Cocaine abuse, in remission
Substance-use -disorders	F14.120	ICD-10	Cocaine abuse with intoxication, uncomplicated
Substance-use -disorders	F14.121	ICD-10	Cocaine abuse with intoxication with delirium
Substance-use -disorders	F14.122	ICD-10	Cocaine abuse with intoxication with perceptual disturbance
Substance-use -disorders	F14.129	ICD-10	Cocaine abuse with intoxication, unspecified
Substance-use -disorders	F14.14	ICD-10	Cocaine abuse with cocaine-induced mood disorder
Substance-use -disorders	F14.150	ICD-10	Cocaine abuse with cocaine-induced psychotic disorder with delusions
Substance-use -disorders	F14.151	ICD-10	Cocaine abuse with cocaine-induced psychotic disorder with hallucinations
Substance-use -disorders	F14.159	ICD-10	Cocaine abuse with cocaine-induced psychotic disorder, unspecified
Substance-use -disorders	F14.180	ICD-10	Cocaine abuse with cocaine-induced anxiety disorder
Substance-use -disorders	F14.181	ICD-10	Cocaine abuse with cocaine-induced sexual dysfunction
Substance-use -disorders	F14.182	ICD-10	Cocaine abuse with cocaine-induced sleep disorder
Substance-use -disorders	F14.188	ICD-10	Cocaine abuse with other cocaine-induced disorder
Substance-use -disorders	F14.19	ICD-10	Cocaine abuse with unspecified cocaine-induced disorder
Substance-use -disorders	F14.20	ICD-10	Cocaine dependence, uncomplicated
Substance-use -disorders	F14.21	ICD-10	Cocaine dependence, in remission
Substance-use -disorders	F14.220	ICD-10	Cocaine dependence with intoxication, uncomplicated
Substance-use -disorders	F14.221	ICD-10	Cocaine dependence with intoxication delirium
Substance-use -disorders	F14.222	ICD-10	Cocaine dependence with intoxication with perceptual disturbance
Substance-use -disorders	F14.229	ICD-10	Cocaine dependence with intoxication, unspecified
Substance-use -disorders	F14.23	ICD-10	Cocaine dependence with withdrawal
Substance-use -disorders	F14.24	ICD-10	Cocaine dependence with cocaine-induced mood disorder

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F14.250	ICD-10	Cocaine dependence with cocaine-induced psychotic disorder with delusions
Substance-use -disorders	F14.251	ICD-10	Cocaine dependence with cocaine-induced psychotic disorder with hallucinations
Substance-use -disorders	F14.259	ICD-10	Cocaine dependence with cocaine-induced psychotic disorder, unspecified
Substance-use -disorders	F14.280	ICD-10	Cocaine dependence with cocaine-induced anxiety disorder
Substance-use -disorders	F14.281	ICD-10	Cocaine dependence with cocaine-induced sexual dysfunction
Substance-use -disorders	F14.282	ICD-10	Cocaine dependence with cocaine-induced sleep disorder
Substance-use -disorders	F14.288	ICD-10	Cocaine dependence with other cocaine-induced disorder
Substance-use -disorders	F14.29	ICD-10	Cocaine dependence with unspecified cocaine-induced disorder
Substance-use -disorders	F14.90	ICD-10	Cocaine use, unspecified, uncomplicated
Substance-use -disorders	F14.920	ICD-10	Cocaine use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F14.921	ICD-10	Cocaine use, unspecified with intoxication delirium
Substance-use -disorders	F14.922	ICD-10	Cocaine use, unspecified with intoxication with perceptual disturbance
Substance-use -disorders	F14.929	ICD-10	Cocaine use, unspecified with intoxication, unspecified
Substance-use -disorders	F14.94	ICD-10	Cocaine use, unspecified with cocaine-induced mood disorder
Substance-use -disorders	F14.950	ICD-10	Cocaine use, unspecified with cocaine-induced psychotic disorder with delusions
Substance-use -disorders	F14.951	ICD-10	Cocaine use, unspecified with cocaine-induced psychotic disorder with hallucinations
Substance-use -disorders	F14.959	ICD-10	Cocaine use, unspecified with cocaine-induced psychotic disorder, unspecified
Substance-use -disorders	F14.980	ICD-10	Cocaine use, unspecified with cocaine-induced anxiety disorder
Substance-use -disorders	F14.981	ICD-10	Cocaine use, unspecified with cocaine-induced sexual dysfunction
Substance-use -disorders	F14.982	ICD-10	Cocaine use, unspecified with cocaine-induced sleep disorder
Substance-use -disorders	F14.988	ICD-10	Cocaine use, unspecified with other cocaine-induced disorder
Substance-use -disorders	F14.99	ICD-10	Cocaine use, unspecified with unspecified cocaine-induced disorder
Substance-use -disorders	F15.10	ICD-10	Other stimulant abuse, uncomplicated
Substance-use -disorders	F15.11	ICD-10	Other stimulant abuse, in remission
Substance-use -disorders	F15.120	ICD-10	Other stimulant abuse with intoxication, uncomplicated

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F15.121	ICD-10	Other stimulant abuse with intoxication delirium
Substance-use -disorders	F15.122	ICD-10	Other stimulant abuse with intoxication with perceptual disturbance
Substance-use -disorders	F15.129	ICD-10	Other stimulant abuse with intoxication, unspecified
Substance-use -disorders	F15.14	ICD-10	Other stimulant abuse with stimulant-induced mood disorder
Substance-use -disorders	F15.150	ICD-10	Other stimulant abuse with stimulant-induced psychotic disorder with delusions
Substance-use -disorders	F15.151	ICD-10	Other stimulant abuse with stimulant-induced psychotic disorder with hallucinations
Substance-use -disorders	F15.159	ICD-10	Other stimulant abuse with stimulant-induced psychotic disorder, unspecified
Substance-use -disorders	F15.180	ICD-10	Other stimulant abuse with stimulant-induced anxiety disorder
Substance-use -disorders	F15.181	ICD-10	Other stimulant abuse with stimulant-induced sexual dysfunction
Substance-use -disorders	F15.182	ICD-10	Other stimulant abuse with stimulant-induced sleep disorder
Substance-use -disorders	F15.188	ICD-10	Other stimulant abuse with other stimulant-induced disorder
Substance-use -disorders	F15.19	ICD-10	Other stimulant abuse with unspecified stimulant-induced disorder
Substance-use -disorders	F15.20	ICD-10	Other stimulant dependence, uncomplicated
Substance-use -disorders	F15.21	ICD-10	Other stimulant dependence, in remission
Substance-use -disorders	F15.220	ICD-10	Other stimulant dependence with intoxication, uncomplicated
Substance-use -disorders	F15.221	ICD-10	Other stimulant dependence with intoxication delirium
Substance-use -disorders	F15.222	ICD-10	Other stimulant dependence with intoxication with perceptual disturbance
Substance-use -disorders	F15.229	ICD-10	Other stimulant dependence with intoxication, unspecified
Substance-use -disorders	F15.23	ICD-10	Other stimulant dependence with withdrawal
Substance-use -disorders	F15.24	ICD-10	Other stimulant dependence with stimulant-induced mood disorder
Substance-use -disorders	F15.250	ICD-10	Other stimulant dependence with stimulant-induced psychotic disorder with delusions
Substance-use -disorders	F15.251	ICD-10	Other stimulant dependence with stimulant-induced psychotic disorder with hallucinations
Substance-use -disorders	F15.259	ICD-10	Other stimulant dependence with stimulant-induced psychotic disorder, unspecified
Substance-use -disorders	F15.280	ICD-10	Other stimulant dependence with stimulant-induced anxiety disorder
Substance-use -disorders	F15.281	ICD-10	Other stimulant dependence with stimulant-induced sexual dysfunction

For peer review only

Online supplementary material

Substance-use -disorders	F15.282	ICD-10	Other stimulant dependence with stimulant-induced sleep disorder
Substance-use -disorders	F15.288	ICD-10	Other stimulant dependence with other stimulant-induced disorder
Substance-use -disorders	F15.29	ICD-10	Other stimulant dependence with unspecified stimulant-induced disorder
Substance-use -disorders	F15.90	ICD-10	Other stimulant use, unspecified, uncomplicated
Substance-use -disorders	F15.920	ICD-10	Other stimulant use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F15.921	ICD-10	Other stimulant use, unspecified with intoxication delirium
Substance-use -disorders	F15.922	ICD-10	Other stimulant use, unspecified with intoxication with perceptual disturbance
Substance-use -disorders	F15.929	ICD-10	Other stimulant use, unspecified with intoxication, unspecified
Substance-use -disorders	F15.93	ICD-10	Other stimulant use, unspecified with withdrawal
Substance-use -disorders	F15.94	ICD-10	Other stimulant use, unspecified with stimulant-induced mood disorder
Substance-use -disorders	F15.950	ICD-10	Other stimulant use, unspecified with stimulant-induced psychotic disorder with delusions
Substance-use -disorders	F15.951	ICD-10	Other stimulant use, unspecified with stimulant-induced psychotic disorder with hallucinations
Substance-use -disorders	F15.959	ICD-10	Other stimulant use, unspecified with stimulant-induced psychotic disorder, unspecified
Substance-use -disorders	F15.980	ICD-10	Other stimulant use, unspecified with stimulant-induced anxiety disorder
Substance-use -disorders	F15.981	ICD-10	Other stimulant use, unspecified with stimulant-induced sexual dysfunction
Substance-use -disorders	F15.982	ICD-10	Other stimulant use, unspecified with stimulant-induced sleep disorder
Substance-use -disorders	F15.988	ICD-10	Other stimulant use, unspecified with other stimulant-induced disorder
Substance-use -disorders	F15.99	ICD-10	Other stimulant use, unspecified with unspecified stimulant-induced disorder
Substance-use -disorders	F16.10	ICD-10	Hallucinogen abuse, uncomplicated
Substance-use -disorders	F16.11	ICD-10	Hallucinogen abuse, in remission
Substance-use -disorders	F16.120	ICD-10	Hallucinogen abuse with intoxication, uncomplicated
Substance-use -disorders	F16.121	ICD-10	Hallucinogen abuse with intoxication with delirium
Substance-use -disorders	F16.122	ICD-10	Hallucinogen abuse with intoxication with perceptual disturbance
Substance-use -disorders	F16.129	ICD-10	Hallucinogen abuse with intoxication, unspecified
Substance-use -disorders	F16.14	ICD-10	Hallucinogen abuse with hallucinogen-induced mood disorder

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F16.150	ICD-10	Hallucinogen abuse with hallucinogen-induced psychotic disorder with delusions
Substance-use -disorders	F16.151	ICD-10	Hallucinogen abuse with hallucinogen-induced psychotic disorder with hallucinations
Substance-use -disorders	F16.159	ICD-10	Hallucinogen abuse with hallucinogen-induced psychotic disorder, unspecified
Substance-use -disorders	F16.180	ICD-10	Hallucinogen abuse with hallucinogen-induced anxiety disorder
Substance-use -disorders	F16.183	ICD-10	Hallucinogen abuse with hallucinogen persisting perception disorder (flashbacks)
Substance-use -disorders	F16.188	ICD-10	Hallucinogen abuse with other hallucinogen-induced disorder
Substance-use -disorders	F16.19	ICD-10	Hallucinogen abuse with unspecified hallucinogen-induced disorder
Substance-use -disorders	F16.20	ICD-10	Hallucinogen dependence, uncomplicated
Substance-use -disorders	F16.21	ICD-10	Hallucinogen dependence, in remission
Substance-use -disorders	F16.220	ICD-10	Hallucinogen dependence with intoxication, uncomplicated
Substance-use -disorders	F16.221	ICD-10	Hallucinogen dependence with intoxication with delirium
Substance-use -disorders	F16.229	ICD-10	Hallucinogen dependence with intoxication, unspecified
Substance-use -disorders	F16.24	ICD-10	Hallucinogen dependence with hallucinogen-induced mood disorder
Substance-use -disorders	F16.250	ICD-10	Hallucinogen dependence with hallucinogen-induced psychotic disorder with delusions
Substance-use -disorders	F16.251	ICD-10	Hallucinogen dependence with hallucinogen-induced psychotic disorder with hallucinations
Substance-use -disorders	F16.259	ICD-10	Hallucinogen dependence with hallucinogen-induced psychotic disorder, unspecified
Substance-use -disorders	F16.280	ICD-10	Hallucinogen dependence with hallucinogen-induced anxiety disorder
Substance-use -disorders	F16.283	ICD-10	Hallucinogen dependence with hallucinogen persisting perception disorder (flashbacks)
Substance-use -disorders	F16.288	ICD-10	Hallucinogen dependence with other hallucinogen-induced disorder
Substance-use -disorders	F16.29	ICD-10	Hallucinogen dependence with unspecified hallucinogen-induced disorder
Substance-use -disorders	F16.90	ICD-10	Hallucinogen use, unspecified, uncomplicated
Substance-use -disorders	F16.920	ICD-10	Hallucinogen use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F16.921	ICD-10	Hallucinogen use, unspecified with intoxication with delirium
Substance-use -disorders	F16.929	ICD-10	Hallucinogen use, unspecified with intoxication, unspecified
Substance-use -disorders	F16.94	ICD-10	Hallucinogen use, unspecified with hallucinogen-induced mood disorder

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F16.950	ICD-10	Hallucinogen use, unspecified with hallucinogen-induced psychotic disorder with delusions
Substance-use -disorders	F16.951	ICD-10	Hallucinogen use, unspecified with hallucinogen-induced psychotic disorder with hallucinations
Substance-use -disorders	F16.959	ICD-10	Hallucinogen use, unspecified with hallucinogen-induced psychotic disorder, unspecified
Substance-use -disorders	F16.980	ICD-10	Hallucinogen use, unspecified with hallucinogen-induced anxiety disorder
Substance-use -disorders	F16.983	ICD-10	Hallucinogen use, unspecified with hallucinogen persisting perception disorder (flashbacks)
Substance-use -disorders	F16.988	ICD-10	Hallucinogen use, unspecified with other hallucinogen-induced disorder
Substance-use -disorders	F16.99	ICD-10	Hallucinogen use, unspecified with unspecified hallucinogen-induced disorder
Substance-use -disorders	F17.200	ICD-10	Nicotine dependence, unspecified, uncomplicated
Substance-use -disorders	F17.201	ICD-10	Nicotine dependence, unspecified, in remission
Substance-use -disorders	F17.203	ICD-10	Nicotine dependence unspecified, with withdrawal
Substance-use -disorders	F17.208	ICD-10	Nicotine dependence, unspecified, with other nicotine-induced disorders
Substance-use -disorders	F17.209	ICD-10	Nicotine dependence, unspecified, with unspecified nicotine-induced disorders
Substance-use -disorders	F17.210	ICD-10	Nicotine dependence, cigarettes, uncomplicated
Substance-use -disorders	F17.211	ICD-10	Nicotine dependence, cigarettes, in remission
Substance-use -disorders	F17.213	ICD-10	Nicotine dependence, cigarettes, with withdrawal
Substance-use -disorders	F17.218	ICD-10	Nicotine dependence, cigarettes, with other nicotine-induced disorders
Substance-use -disorders	F17.219	ICD-10	Nicotine dependence, cigarettes, with unspecified nicotine-induced disorders
Substance-use -disorders	F17.220	ICD-10	Nicotine dependence, chewing tobacco, uncomplicated
Substance-use -disorders	F17.221	ICD-10	Nicotine dependence, chewing tobacco, in remission
Substance-use -disorders	F17.223	ICD-10	Nicotine dependence, chewing tobacco, with withdrawal
Substance-use -disorders	F17.228	ICD-10	Nicotine dependence, chewing tobacco, with other nicotine-induced disorders
Substance-use -disorders	F17.229	ICD-10	Nicotine dependence, chewing tobacco, with unspecified nicotine-induced disorders
Substance-use -disorders	F17.290	ICD-10	Nicotine dependence, other tobacco product, uncomplicated
Substance-use -disorders	F17.291	ICD-10	Nicotine dependence, other tobacco product, in remission
Substance-use -disorders	F17.293	ICD-10	Nicotine dependence, other tobacco product, with withdrawal

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F17.298	ICD-10	Nicotine dependence, other tobacco product, with other nicotine-induced disorders
Substance-use -disorders	F17.299	ICD-10	Nicotine dependence, other tobacco product, with unspecified nicotine-induced disorders
Substance-use -disorders	F18.10	ICD-10	Inhalant abuse, uncomplicated
Substance-use -disorders	F18.11	ICD-10	Inhalant abuse, in remission
Substance-use -disorders	F18.120	ICD-10	Inhalant abuse with intoxication, uncomplicated
Substance-use -disorders	F18.121	ICD-10	Inhalant abuse with intoxication delirium
Substance-use -disorders	F18.129	ICD-10	Inhalant abuse with intoxication, unspecified
Substance-use -disorders	F18.14	ICD-10	Inhalant abuse with inhalant-induced mood disorder
Substance-use -disorders	F18.150	ICD-10	Inhalant abuse with inhalant-induced psychotic disorder with delusions
Substance-use -disorders	F18.151	ICD-10	Inhalant abuse with inhalant-induced psychotic disorder with hallucinations
Substance-use -disorders	F18.159	ICD-10	Inhalant abuse with inhalant-induced psychotic disorder, unspecified
Substance-use -disorders	F18.17	ICD-10	Inhalant abuse with inhalant-induced dementia
Substance-use -disorders	F18.180	ICD-10	Inhalant abuse with inhalant-induced anxiety disorder
Substance-use -disorders	F18.188	ICD-10	Inhalant abuse with other inhalant-induced disorder
Substance-use -disorders	F18.19	ICD-10	Inhalant abuse with unspecified inhalant-induced disorder
Substance-use -disorders	F18.20	ICD-10	Inhalant dependence, uncomplicated
Substance-use -disorders	F18.21	ICD-10	Inhalant dependence, in remission
Substance-use -disorders	F18.220	ICD-10	Inhalant dependence with intoxication, uncomplicated
Substance-use -disorders	F18.221	ICD-10	Inhalant dependence with intoxication delirium
Substance-use -disorders	F18.229	ICD-10	Inhalant dependence with intoxication, unspecified
Substance-use -disorders	F18.24	ICD-10	Inhalant dependence with inhalant-induced mood disorder
Substance-use -disorders	F18.250	ICD-10	Inhalant dependence with inhalant-induced psychotic disorder with delusions
Substance-use -disorders	F18.251	ICD-10	Inhalant dependence with inhalant-induced psychotic disorder with hallucinations
Substance-use -disorders	F18.259	ICD-10	Inhalant dependence with inhalant-induced psychotic disorder, unspecified
Substance-use -disorders	F18.27	ICD-10	Inhalant dependence with inhalant-induced dementia

For peer review only

Online supplementary material

Substance-use -disorders	F18.280	ICD-10	Inhalant dependence with inhalant-induced anxiety disorder
Substance-use -disorders	F18.288	ICD-10	Inhalant dependence with other inhalant-induced disorder
Substance-use -disorders	F18.29	ICD-10	Inhalant dependence with unspecified inhalant-induced disorder
Substance-use -disorders	F18.90	ICD-10	Inhalant use, unspecified, uncomplicated
Substance-use -disorders	F18.920	ICD-10	Inhalant use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F18.921	ICD-10	Inhalant use, unspecified with intoxication with delirium
Substance-use -disorders	F18.929	ICD-10	Inhalant use, unspecified with intoxication, unspecified
Substance-use -disorders	F18.94	ICD-10	Inhalant use, unspecified with inhalant-induced mood disorder
Substance-use -disorders	F18.950	ICD-10	Inhalant use, unspecified with inhalant-induced psychotic disorder with delusions
Substance-use -disorders	F18.951	ICD-10	Inhalant use, unspecified with inhalant-induced psychotic disorder with hallucinations
Substance-use -disorders	F18.959	ICD-10	Inhalant use, unspecified with inhalant-induced psychotic disorder, unspecified
Substance-use -disorders	F18.97	ICD-10	Inhalant use, unspecified with inhalant-induced persisting dementia
Substance-use -disorders	F18.980	ICD-10	Inhalant use, unspecified with inhalant-induced anxiety disorder
Substance-use -disorders	F18.988	ICD-10	Inhalant use, unspecified with other inhalant-induced disorder
Substance-use -disorders	F18.99	ICD-10	Inhalant use, unspecified with unspecified inhalant-induced disorder
Substance-use -disorders	F19.10	ICD-10	Other psychoactive substance abuse, uncomplicated
Substance-use -disorders	F19.11	ICD-10	Other psychoactive substance abuse, in remission
Substance-use -disorders	F19.120	ICD-10	Other psychoactive substance abuse with intoxication, uncomplicated
Substance-use -disorders	F19.121	ICD-10	Other psychoactive substance abuse with intoxication delirium
Substance-use -disorders	F19.122	ICD-10	Other psychoactive substance abuse with intoxication with perceptual disturbances
Substance-use -disorders	F19.129	ICD-10	Other psychoactive substance abuse with intoxication, unspecified
Substance-use -disorders	F19.14	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced mood disorder
Substance-use -disorders	F19.150	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced psychotic disorder with delusions
Substance-use -disorders	F19.151	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced psychotic disorder with hallucinations
Substance-use -disorders	F19.159	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced psychotic disorder, unspecified

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F19.16	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced persisting amnestic disorder
Substance-use -disorders	F19.17	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced persisting dementia
Substance-use -disorders	F19.180	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced anxiety disorder
Substance-use -disorders	F19.181	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced sexual dysfunction
Substance-use -disorders	F19.182	ICD-10	Other psychoactive substance abuse with psychoactive substance-induced sleep disorder
Substance-use -disorders	F19.188	ICD-10	Other psychoactive substance abuse with other psychoactive substance-induced disorder
Substance-use -disorders	F19.19	ICD-10	Other psychoactive substance abuse with unspecified psychoactive substance-induced disorder
Substance-use -disorders	F19.20	ICD-10	Other psychoactive substance dependence, uncomplicated
Substance-use -disorders	F19.21	ICD-10	Other psychoactive substance dependence, in remission
Substance-use -disorders	F19.220	ICD-10	Other psychoactive substance dependence with intoxication, uncomplicated
Substance-use -disorders	F19.221	ICD-10	Other psychoactive substance dependence with intoxication delirium
Substance-use -disorders	F19.222	ICD-10	Other psychoactive substance dependence with intoxication with perceptual disturbance
Substance-use -disorders	F19.229	ICD-10	Other psychoactive substance dependence with intoxication, unspecified
Substance-use -disorders	F19.230	ICD-10	Other psychoactive substance dependence with withdrawal, uncomplicated
Substance-use -disorders	F19.231	ICD-10	Other psychoactive substance dependence with withdrawal delirium
Substance-use -disorders	F19.232	ICD-10	Other psychoactive substance dependence with withdrawal with perceptual disturbance
Substance-use -disorders	F19.239	ICD-10	Other psychoactive substance dependence with withdrawal, unspecified
Substance-use -disorders	F19.24	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced mood disorder
Substance-use -disorders	F19.250	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced psychotic disorder with delusions
Substance-use -disorders	F19.251	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced psychotic disorder with hallucinations
Substance-use -disorders	F19.259	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced psychotic disorder, unspecified
Substance-use -disorders	F19.26	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced persisting amnestic disorder
Substance-use -disorders	F19.27	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced persisting dementia

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F19.280	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced anxiety disorder
Substance-use -disorders	F19.281	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced sexual dysfunction
Substance-use -disorders	F19.282	ICD-10	Other psychoactive substance dependence with psychoactive substance-induced sleep disorder
Substance-use -disorders	F19.288	ICD-10	Other psychoactive substance dependence with other psychoactive substance-induced disorder
Substance-use -disorders	F19.29	ICD-10	Other psychoactive substance dependence with unspecified psychoactive substance-induced disorder
Substance-use -disorders	F19.90	ICD-10	Other psychoactive substance use, unspecified, uncomplicated
Substance-use -disorders	F19.920	ICD-10	Other psychoactive substance use, unspecified with intoxication, uncomplicated
Substance-use -disorders	F19.921	ICD-10	Other psychoactive substance use, unspecified with intoxication with delirium
Substance-use -disorders	F19.922	ICD-10	Other psychoactive substance use, unspecified with intoxication with perceptual disturbance
Substance-use -disorders	F19.929	ICD-10	Other psychoactive substance use, unspecified with intoxication, unspecified
Substance-use -disorders	F19.930	ICD-10	Other psychoactive substance use, unspecified with withdrawal, uncomplicated
Substance-use -disorders	F19.931	ICD-10	Other psychoactive substance use, unspecified with withdrawal delirium
Substance-use -disorders	F19.932	ICD-10	Other psychoactive substance use, unspecified with withdrawal with perceptual disturbance
Substance-use -disorders	F19.939	ICD-10	Other psychoactive substance use, unspecified with withdrawal, unspecified
Substance-use -disorders	F19.94	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced mood disorder
Substance-use -disorders	F19.950	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced psychotic disorder with delusions
Substance-use -disorders	F19.951	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced psychotic disorder with hallucinations
Substance-use -disorders	F19.959	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced psychotic disorder, unspecified
Substance-use -disorders	F19.96	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced persisting amnestic disorder
Substance-use -disorders	F19.97	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced persisting dementia
Substance-use -disorders	F19.980	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced anxiety disorder
Substance-use -disorders	F19.981	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced sexual dysfunction
Substance-use -disorders	F19.982	ICD-10	Other psychoactive substance use, unspecified with psychoactive substance-induced sleep disorder

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

Substance-use -disorders	F19.988	ICD-10	Other psychoactive substance use, unspecified with other psychoactive substance-induced disorder
Substance-use -disorders	F19.99	ICD-10	Other psychoactive substance use, unspecified with unspecified psychoactive substance-induced disorder

For peer review only

Online supplementary material

eTable 7: U.S. State of Residence

State of residence	Full name	Number of patients	Percentage
CO	Colorado	125,267	23.0%
NY	New York	103,016	18.9%
MO	Missouri	98,450	18.1%
NC	North Carolina	75,626	13.9%
IN	Indiana	52,699	9.7%
VA	Virginia	23,848	4.4%
KY	Kentucky	18,102	3.3%
VT	Vermont	8,473	1.6%
DE	Delaware	7,313	1.3%
AK	Alaska	4,776	0.9%
MT	Montana	4,696	0.9%
AL	Alabama	3,459	0.6%
NJ	New Jersey	2,074	0.4%
IL	Illinois	1,994	0.4%
SC	South Carolina	1,193	0.2%
CT	Connecticut	1,125	0.2%
KS	Kansas	1,007	0.2%
FL	Florida	932	0.2%
NH	New Hampshire	695	0.1%
CA	California	679	0.1%
GA	Georgia	536	0.1%
PA	Pennsylvania	506	0.1%
TX	Texas	473	0.1%
MD	Maryland	451	0.1%
TN	Tennessee	448	0.1%
OH	Ohio	309	0.1%
WV	West Virginia	297	0.1%
MI	Michigan	272	0.1%
MA	Massachusetts	258	0.0%
WY	Wyoming	213	0.0%
MN	Minnesota	199	0.0%
AZ	Arizona	170	0.0%
WA	Washington	147	0.0%
NM	New Mexico	135	0.0%
WI	Wisconsin	111	0.0%
NE	Nebraska	98	0.0%
OK	Oklahoma	90	0.0%
LA	Louisiana	88	0.0%
IA	Iowa	82	0.0%
AR	Arkansas	78	0.0%

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

OR	Oregon	77	0.0%
SD	South Dakota	73	0.0%
UT	Utah	71	0.0%
DC	District of Columbia	67	0.0%
ND	North Dakota	63	0.0%
NV	Nevada	60	0.0%
MS	Mississippi	53	0.0%
RI	Rhode Island	42	0.0%
ME	Maine	42	0.0%
PR	Puerto Rico	30	0.0%
ID	Idaho	26	0.0%
HI	Hawaii	23	0.0%
AE	Armed Forces Europe	<20	0.0%
AP	Armed Forces Pacific	<20	0.0%

For peer review only

Online supplementary material

eTable 8: Mean CGI-S score by diagnosis

Diagnosis	Number of patients	Mean CGI-S	95% Confidence Interval	Standard deviation
Dementia & Alzheimer's disease	5,029	4.33	4.30 to 4.37	1.36
Substance-use disorders	152,790	3.61	3.60 to 3.62	1.80
Schizophrenia	14,592	4.44	4.41 to 4.46	1.44
Schizoaffective disorder	15,044	4.48	4.45 to 4.50	1.42
Bipolar disorder	69,607	4.28	4.27 to 4.29	1.45
Major depressive disorder (MDD)	129,120	4.21	4.20 to 4.22	1.42
Anxiety disorders	103,923	4.02	4.01 to 4.03	1.53
Attention-deficit hyperactivity disorder (ADHD)	53,744	3.99	3.97 to 4.00	1.42
Note: Confidence intervals were estimated using the formula mean +/- error margin where: Error margin = std/sqrt(N) * Z, Z = 1.96 for 95% confidence interval				

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

eTable 9: Mean maximum and minimum CGI-S score by diagnosis

Diagnosis	Number of patients with at least 2 CGI-S data points	Max CGI-S Mean	Min CGI-S Mean	Max CGI-S Std dev	Min CGI-S Std dev	Max CGI-S Confidence Interval	Min CGI-S Confidence Interval
Dementia & Alzheimer's disease	2,980	5.10	3.32	1.06	1.56	5.06 to5.14	3.26 to 3.37
Substance-use disorders	107,939	4.85	2.55	1.32	1.62	4.84 to 4.85	2.54 to 2.56
Schizophrenia	12,465	5.41	2.90	1.01	1.55	5.40 to 5.43	2.87 to 2.93
Schizoaffective disorder	13,590	5.53	2.88	0.95	1.55	5.51 to 5.55	2.86 to 2.91
Bipolar disorder	59,222	5.22	3.02	0.98	1.54	5.21 to 5.23	3.00 to 3.03
Major depressive disorder (MDD)	103,933	5.02	3.03	0.98	1.53	5.01 to 5.03	3.03 to 3.04
Anxiety disorders	89,710	5.04	2.80	0.97	1.58	5.03 to 5.04	2.79 to 2.81
Attention-deficit hyperactivity disorder (ADHD)	47,591	4.92	2.75	1.00	1.50	4.91 to 4.93	2.73 to 2.76

Note: Confidence intervals were estimated using the formula mean +/- error margin where:
Error margin = std/sqrt(N) * Z, Z = 1.96 for 95% confidence interval

For peer review only

Online supplementary material

eTable 10: Number of patients with at least one hospital visit by diagnosis

Diagnosis	Number of patients	Number of hospitalized patients	Percentage
Dementia & Alzheimer's disease	5,029	2,625	52.2%
Substance-use disorders	152,790	54,267	35.5%
Schizophrenia	14,592	8,123	55.7%
Schizoaffective disorder	15,044	8,528	56.7%
Bipolar disorder	69,607	25,322	36.4%
Major depressive disorder (MDD)	129,120	40,726	31.5%
Anxiety disorders	103,923	22,455	21.6%
Attention-deficit hyperactivity disorder (ADHD)	53,744	8,587	16.0%
Note: Hospitalization includes both Inpatient and Emergency services			

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

eTable 11: SAV features included in MSE NLP analysis (eTable 12)

Category	Description	SAV number
Delusions	experiencing delusions/abnormal thoughts (grandeur)	sav_1_aopt_1
Delusions	experiencing delusions/abnormal thoughts (persecution)	sav_2_aopt_2
Delusions	experiencing delusions/abnormal thoughts (religious)	sav_3_aopt_3
Delusions	experiencing hallucinations (auditory)	sav_4_aopt_4
Delusions	experiencing hallucinations (visual)	sav_5_aopt_5
Delusions	experiencing hallucinations (tactile)	sav_6_aopt_6
Delusions	History of delusions/abnormal thoughts	sav_7_aopt_7
Delusions	History of hallucinations	sav_8_aopt_8
Delusions	experiencing delusions (not specified)	sav_10_aopt_10
Delusions	experiencing hallucinations (not specified)	sav_11_aopt_11
Delusions	experiencing delusions/abnormal thoughts (paranoia)	sav_13_aopt_13
Delusions	experiencing hallucinations (olfactory)	sav_14_aopt_14
Delusions	responding to internal stimuli	sav_15_aopt_15
Delusions	experiencing delusions/abnormal thoughts (sexual)	sav_16_aopt_16
Mood	"anxious, tense"	sav_158_mood_2
Mood	"depressed, sad, despondent"	sav_159_mood_3
Mood	"irritable, angry"	sav_160_mood_4
Mood	declined	sav_162_mood_6
Mood	"elevated, manic"	sav_163_mood_7
Mood	labile	sav_164_mood_8
Mood	misc. issues	sav_165_mood_9
Mood	suicidal	sav_166_mood_10
Cognition	general issues	sav_65_attn_concentration_2
Cognition	declined	sav_67_attn_concentration_4
Cognition	issues due to MH	sav_69_attn_concentration_6
Cognition	issues due to DD	sav_70_attn_concentration_7
Cognition	varying	sav_71_attn_concentration_8
Cognition	issues with concentration	sav_81_cognition_2
Cognition	issues with fund of knowledge	sav_83_cognition_4
Cognition	issues with attention	sav_85_cognition_6
Cognition	general issues	sav_87_cognition_8
Cognition	some impairment	sav_90_exec_functioning_2
Cognition	much impairment	sav_91_exec_functioning_3
Cognition	declined since last session	sav_93_exec_functioning_5
Cognition	below age level	sav_95_exec_functioning_7
Cognition	impaired (due to DD)	sav_96_exec_functioning_8
Cognition	impaired (due to TBI)	sav_97_exec_functioning_9
Cognition	impaired (due to other MH issue)	sav_98_exec_functioning_10

For peer review only

Online supplementary material

eTable 12: Number of patients with at least one MSE feature by diagnosis

Diagnosis	MSE Category	Number of patients	Number of patients with at least one MSE feature	Percentage
Dementia & Alzheimer's disease	Delusions & hallucinations	5,029	2,147	42.7%
Dementia & Alzheimer's disease	Mood	5,029	2,518	50.1%
Dementia & Alzheimer's disease	Cognition	5,029	2,719	54.1%
Substance-use disorders	Delusions & hallucinations	152,790	37,455	24.5%
Substance-use disorders	Mood	152,790	89,616	58.7%
Substance-use disorders	Cognition	152,790	26,461	17.3%
Schizophrenia	Delusions & hallucinations	14,592	11,513	78.9%
Schizophrenia	Mood	14,592	10,753	73.7%
Schizophrenia	Cognition	14,592	6,418	44.0%
Schizoaffective disorder	Delusions & hallucinations	15,044	11,885	79.0%
Schizoaffective disorder	Mood	15,044	12,278	81.6%
Schizoaffective disorder	Cognition	15,044	6,785	45.1%
Bipolar disorder	Delusions & hallucinations	69,607	33,179	47.7%
Bipolar disorder	Mood	69,607	55,472	79.7%
Bipolar disorder	Cognition	69,607	23,373	33.6%
Major depressive disorder (MDD)	Delusions & hallucinations	129,120	42,566	33.0%
Major depressive disorder (MDD)	Mood	129,120	100,182	77.6%
Major depressive disorder (MDD)	Cognition	129,120	28,986	22.4%
Anxiety disorders	Delusions & hallucinations	103,923	32,059	30.8%
Anxiety disorders	Mood	103,923	76,320	73.4%
Anxiety disorders	Cognition	103,923	25,625	24.7%
Attention-deficit hyperactivity disorder (ADHD)	Delusions & hallucinations	53,744	11,933	22.2%
Attention-deficit hyperactivity disorder (ADHD)	Mood	53,744	31,112	57.9%
Attention-deficit hyperactivity disorder (ADHD)	Cognition	53,744	18,432	34.3%

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Online supplementary material

eFigure 1: MindLinc EHR data de-identification and harmonization pipeline

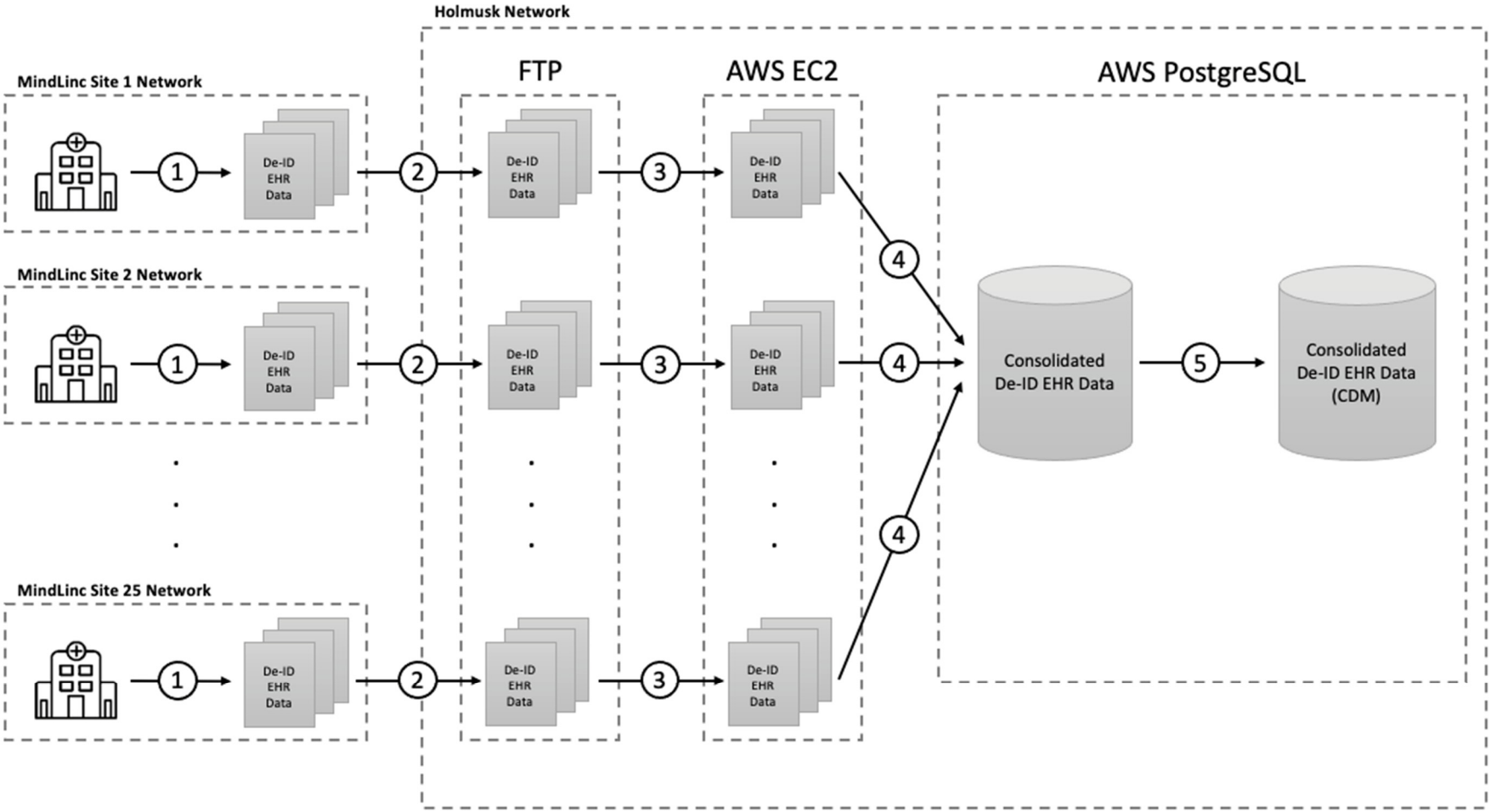


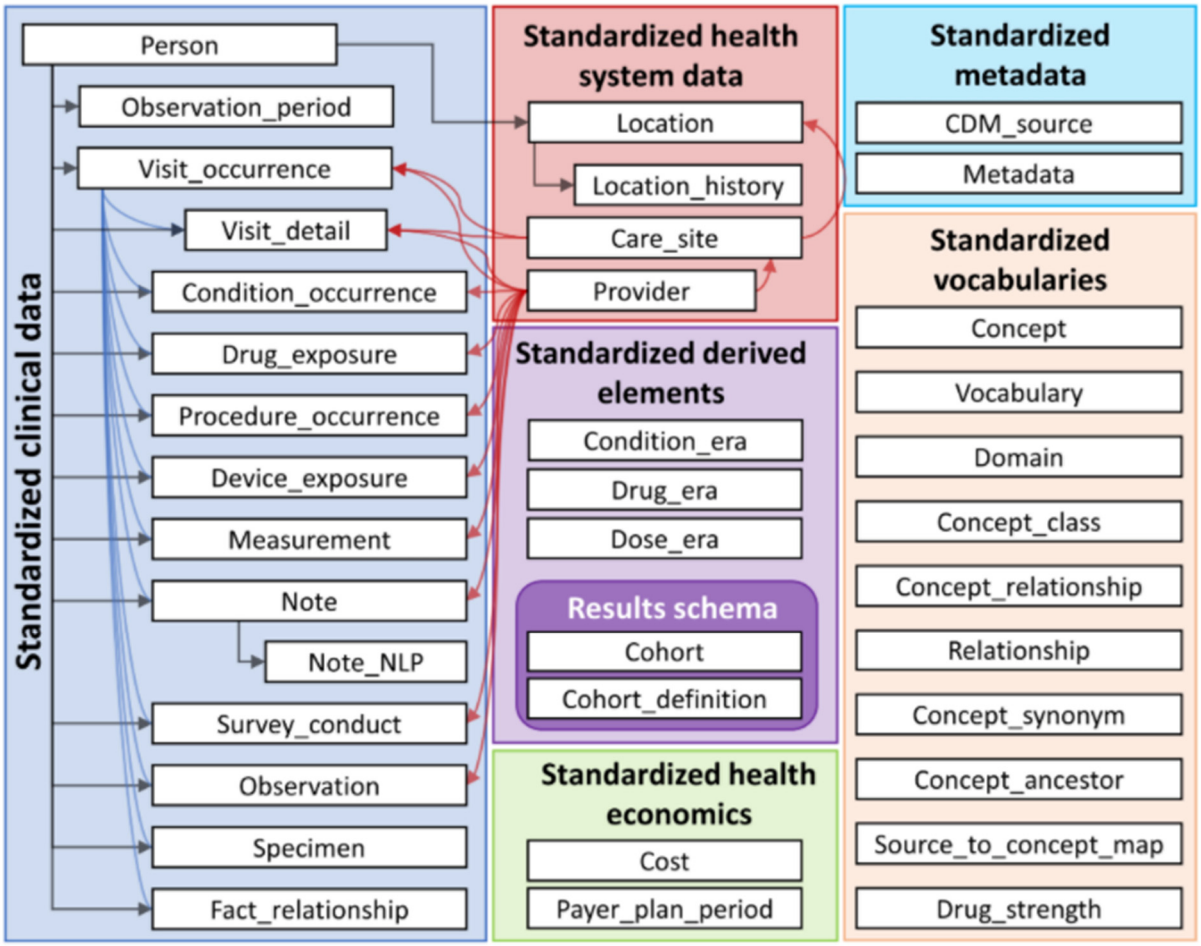
Figure: MindLinc Data Flow Diagram

Online supplementary material

eFigure 2: OMOP CDM schematic

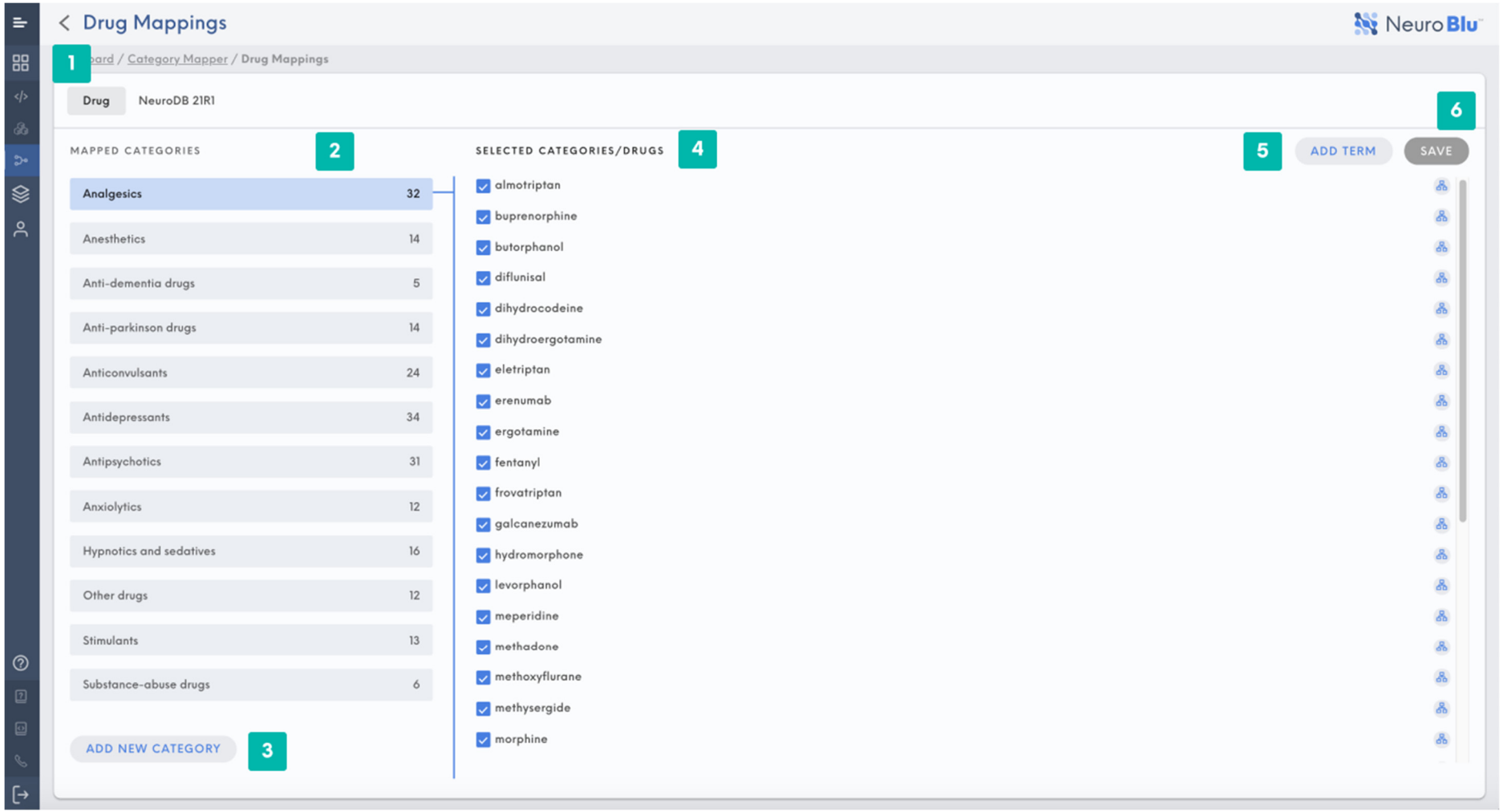
Source: https://ohdsi.github.io/TheBookOfOhdsi/CommonDataModel.html

Attribution: Creative Commons Zero v1.0 Universal license (No copyright)



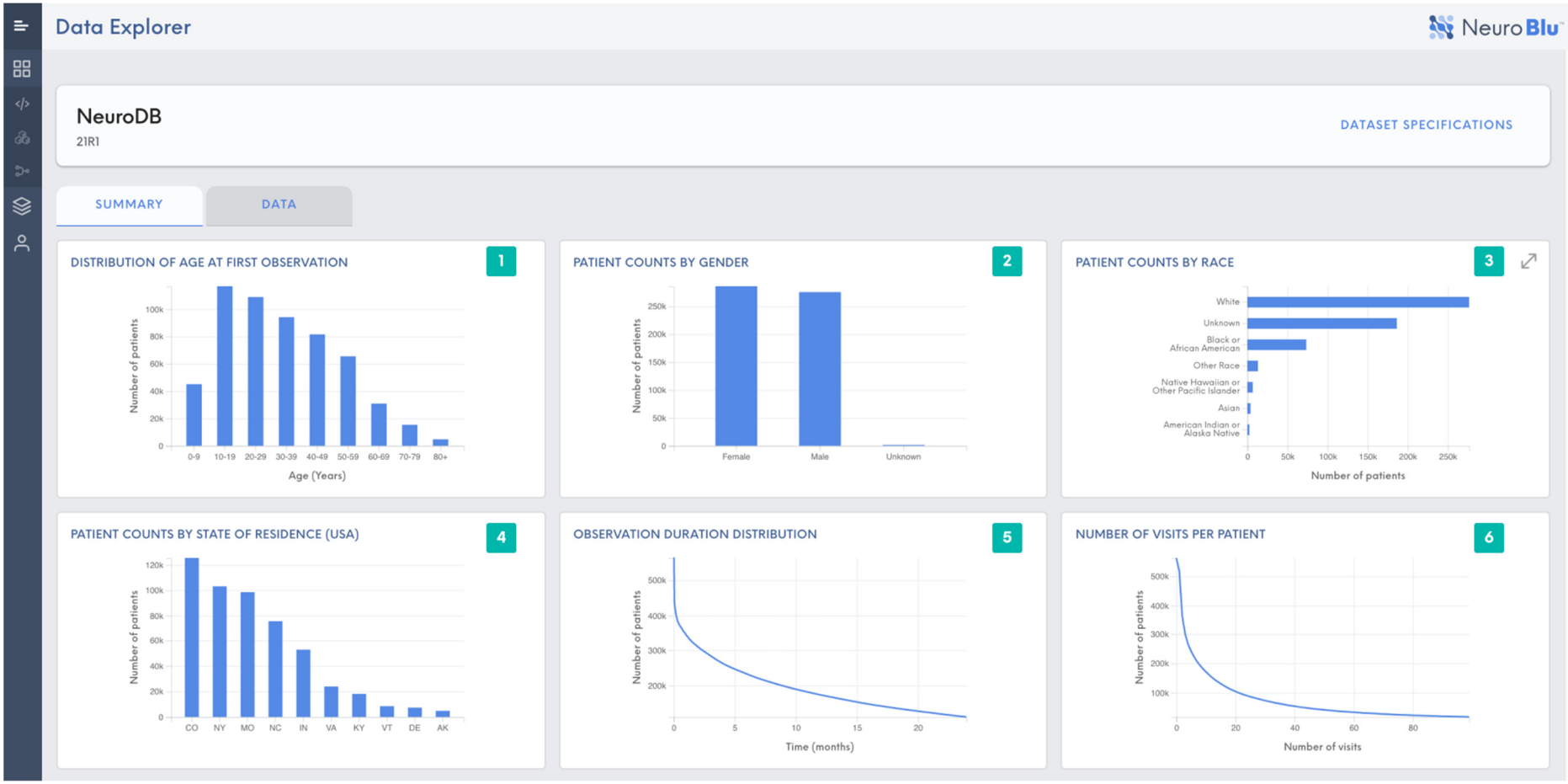
Online supplementary material

eFigure 3: NeuroBlu Category Mapper illustrating the classification of individual drugs into different types of medication

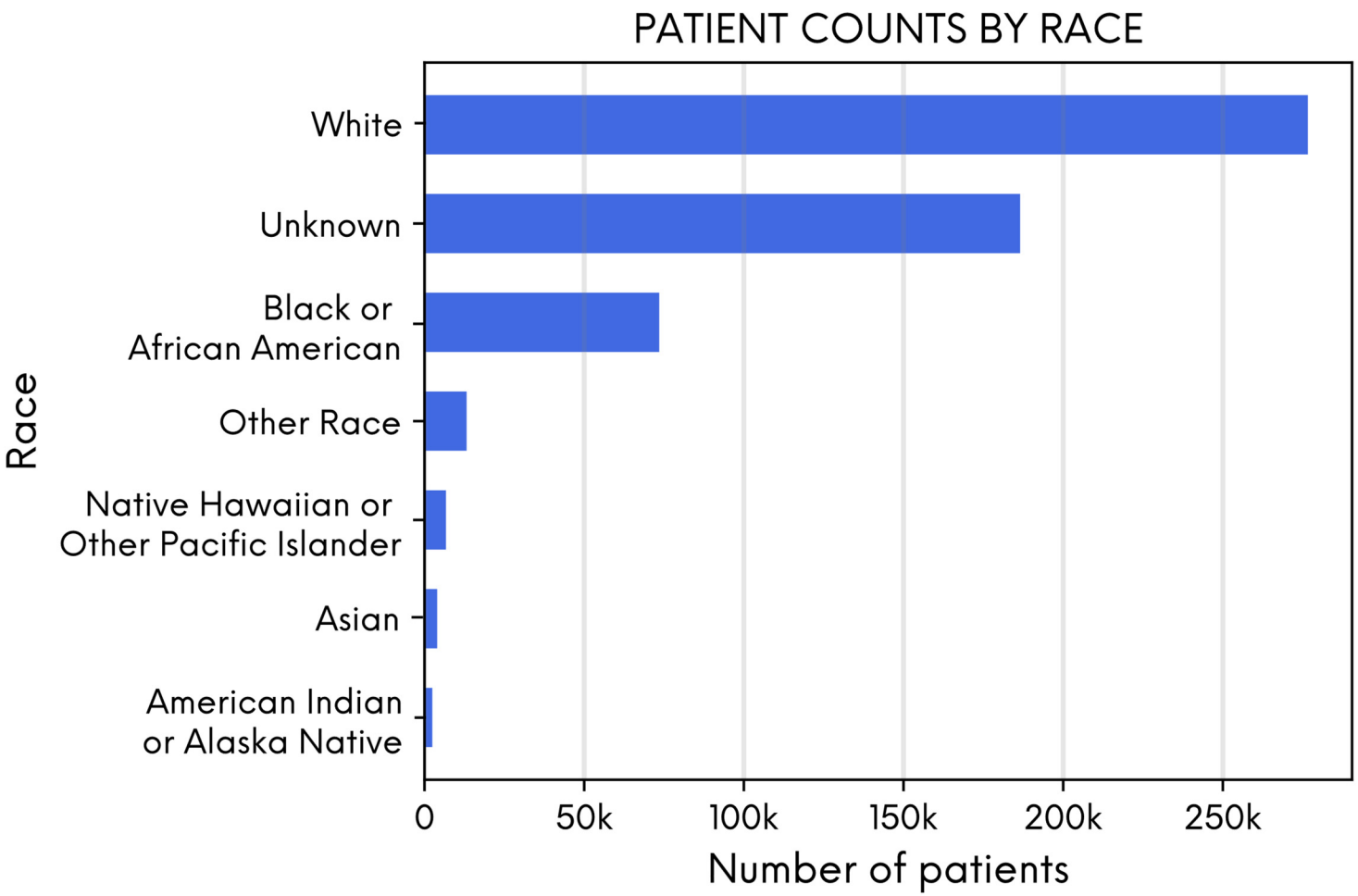


Online supplementary material

eFigure 4: NeuroBlu Data Explorer which provides a summary of the dataset and allows the top 50 rows of each table to be viewed

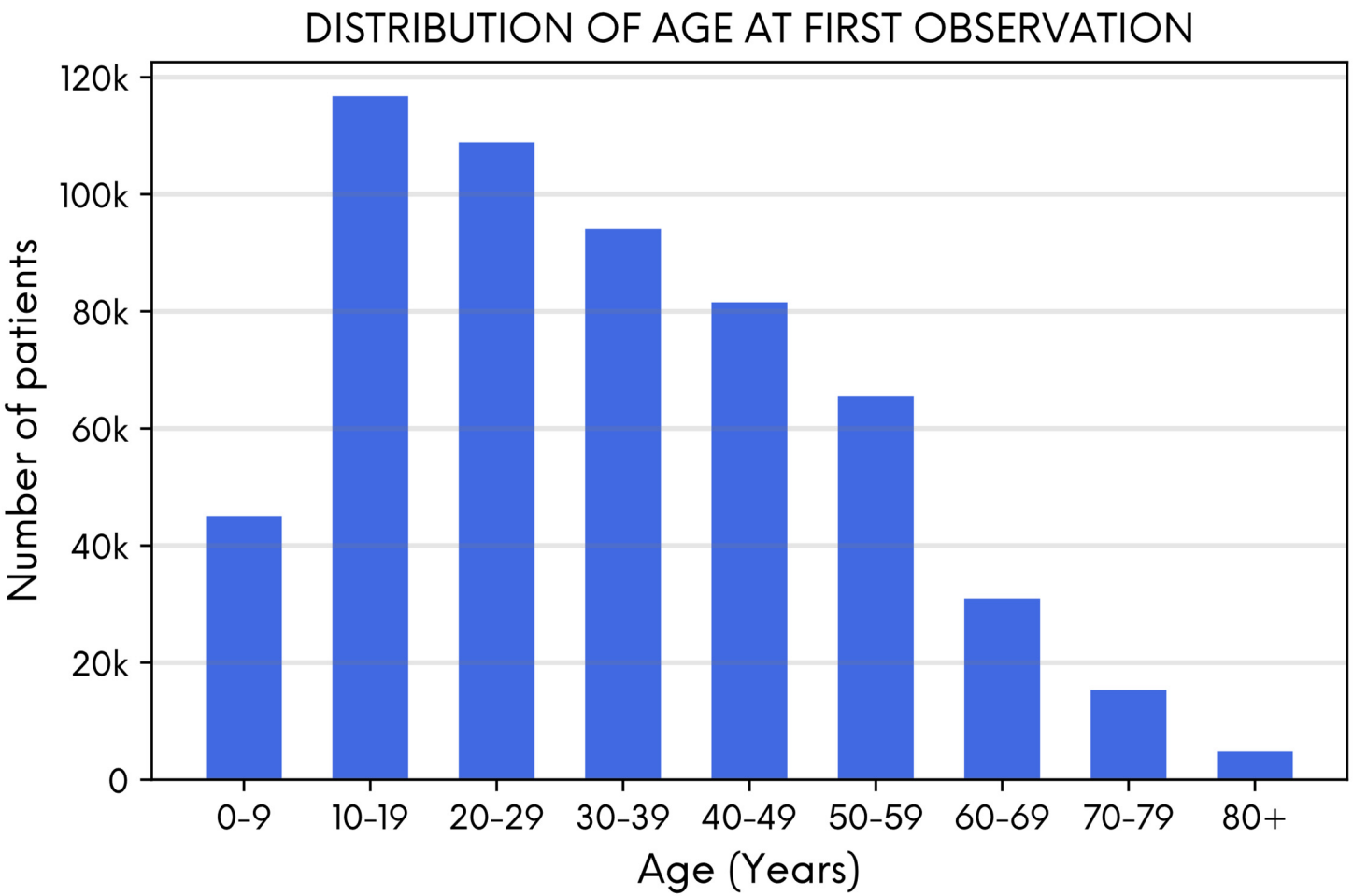


eFigure 5: NeuroBlu dataset race distribution

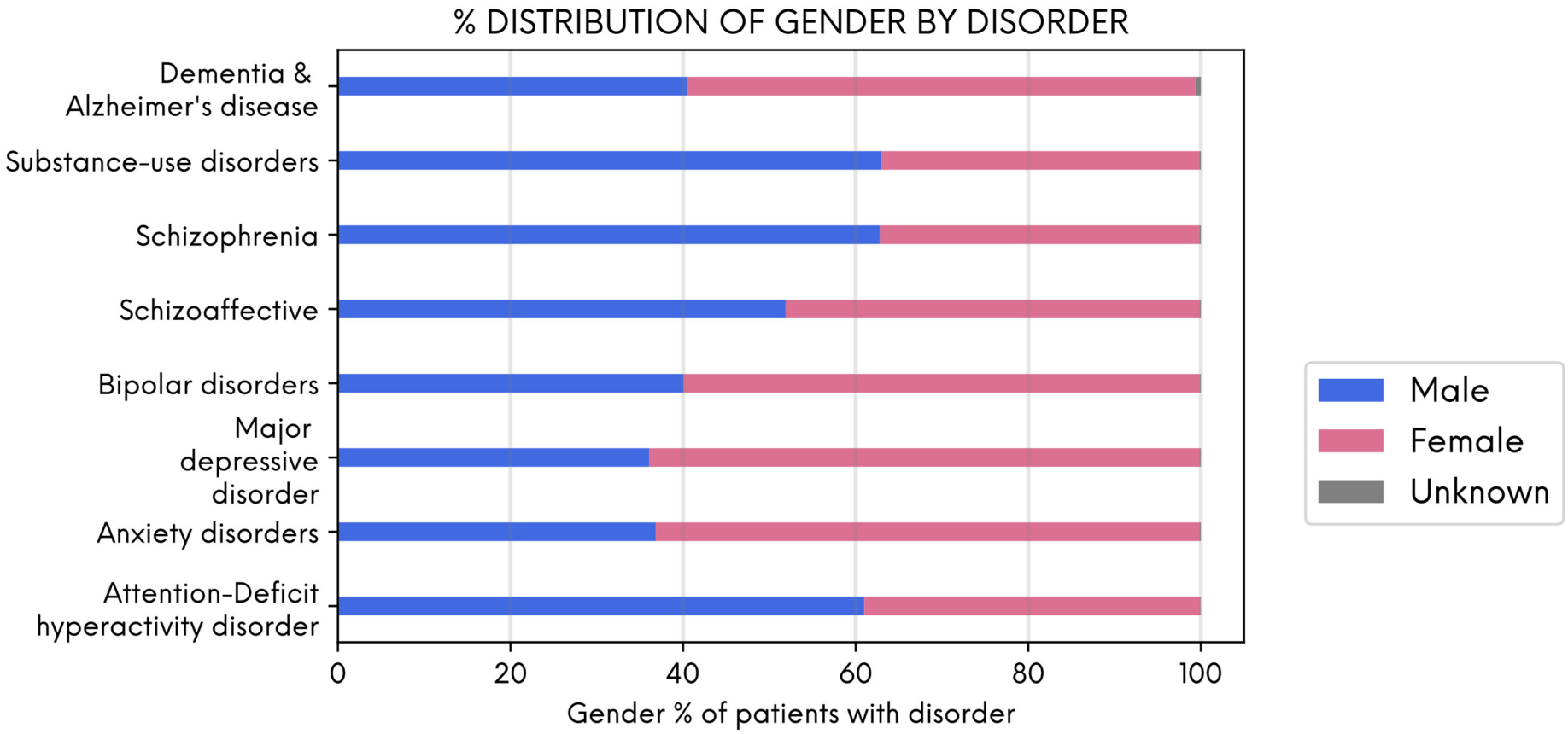


Online supplementary material

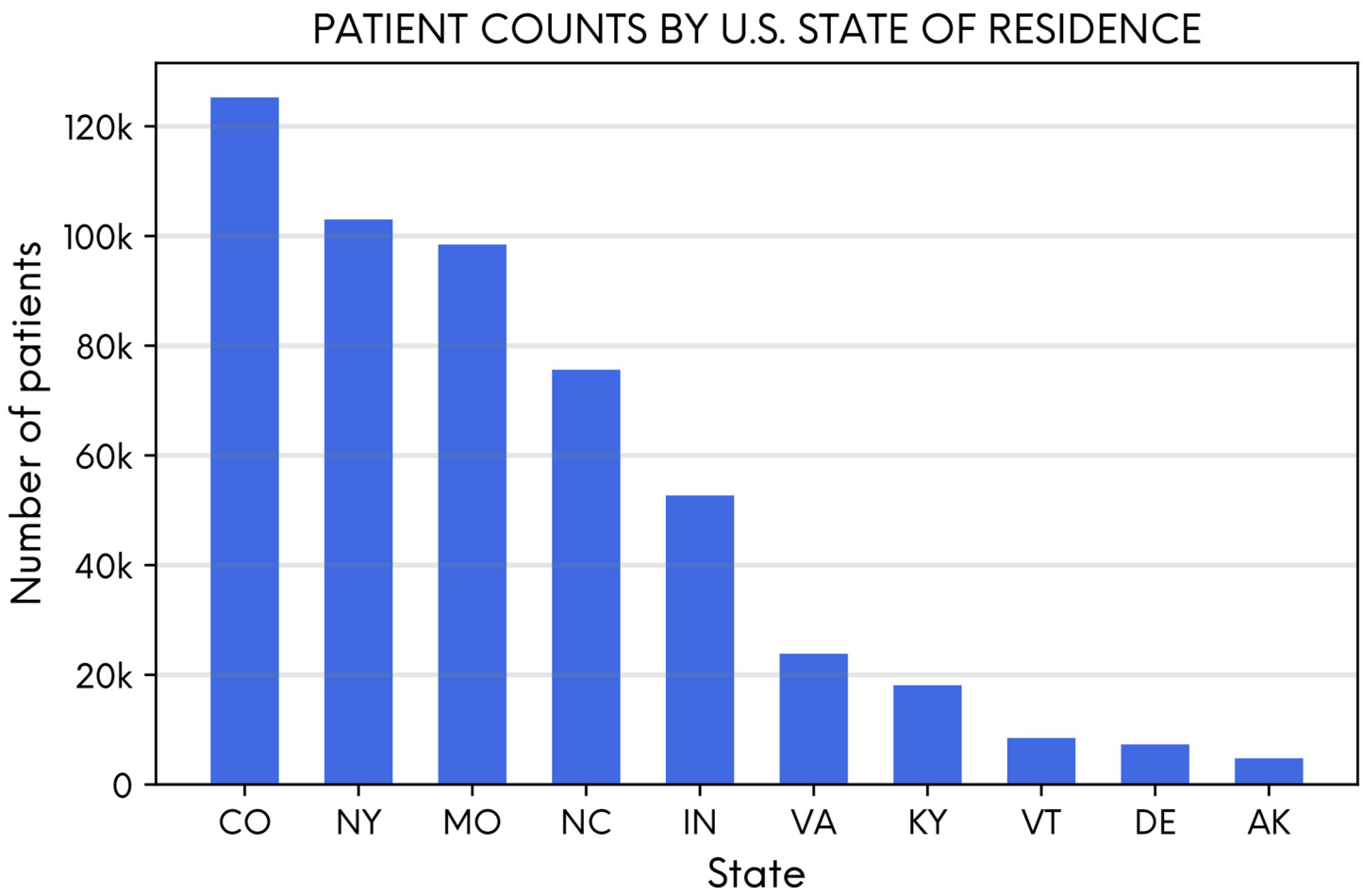
eFigure 6: NeuroBlu dataset age distribution



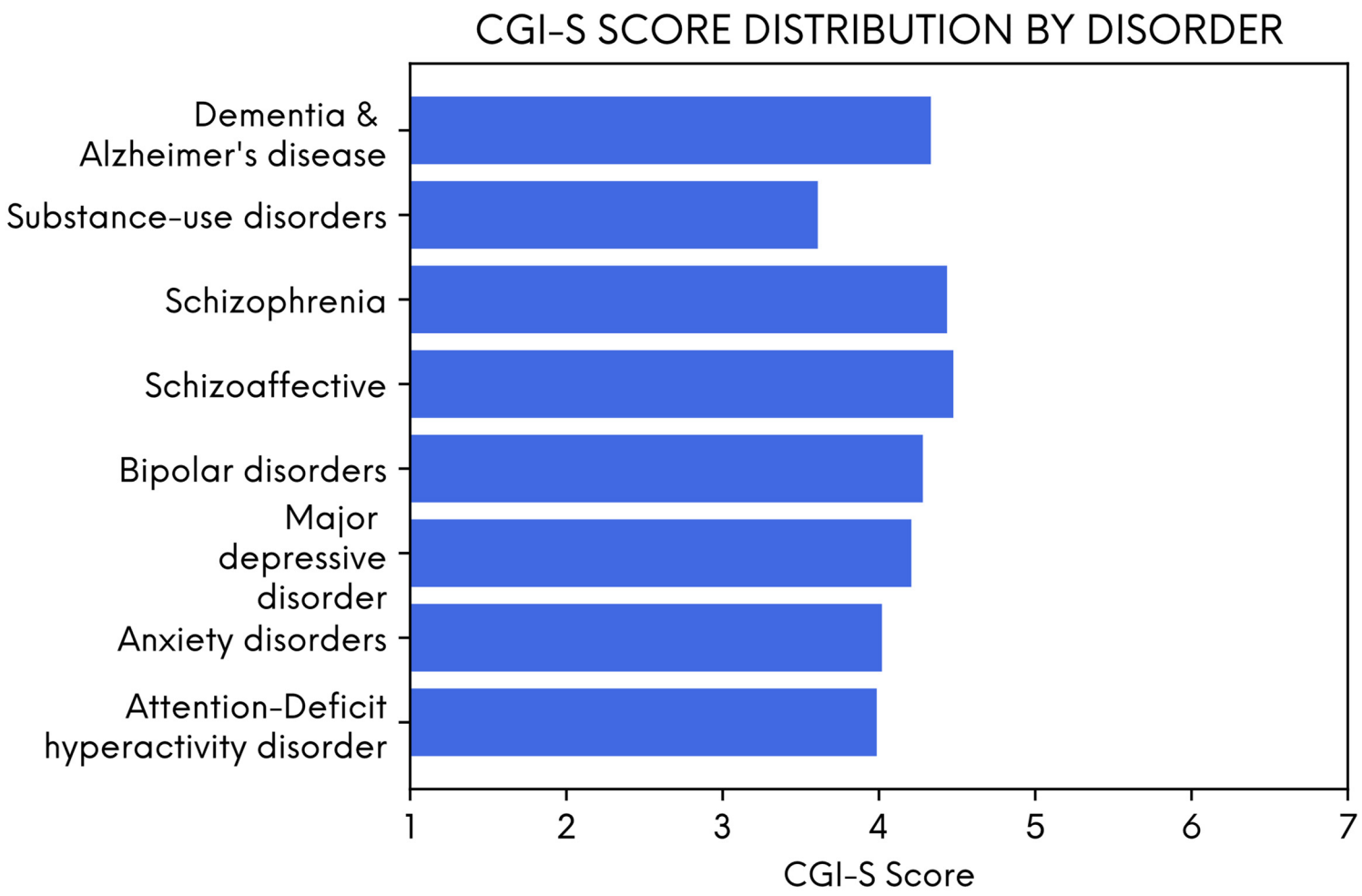
eFigure 7: Gender distribution by diagnosis



eFigure 8: U.S. State of Residence

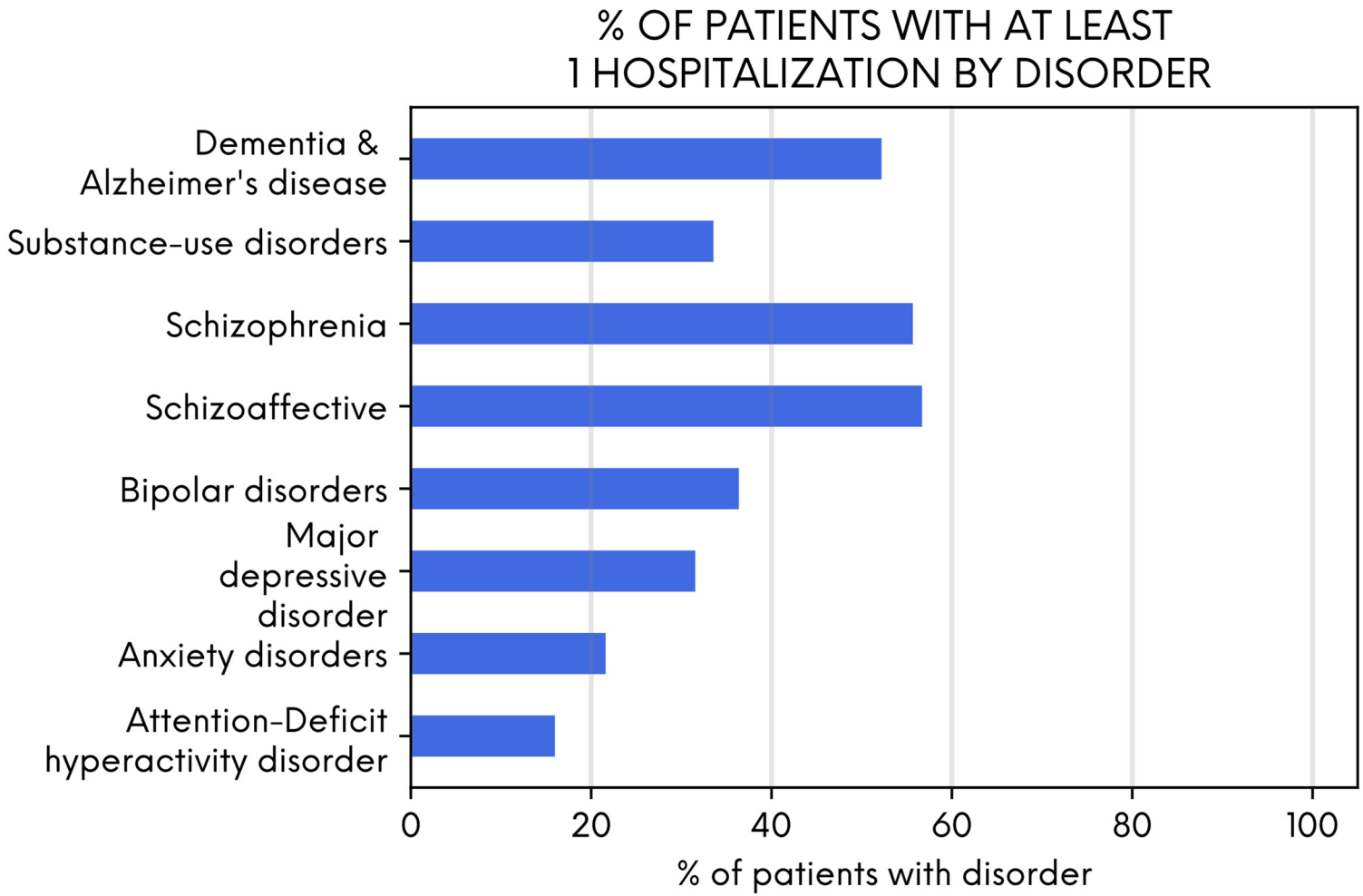


eFigure 9: Mean CGI-S score by diagnosis

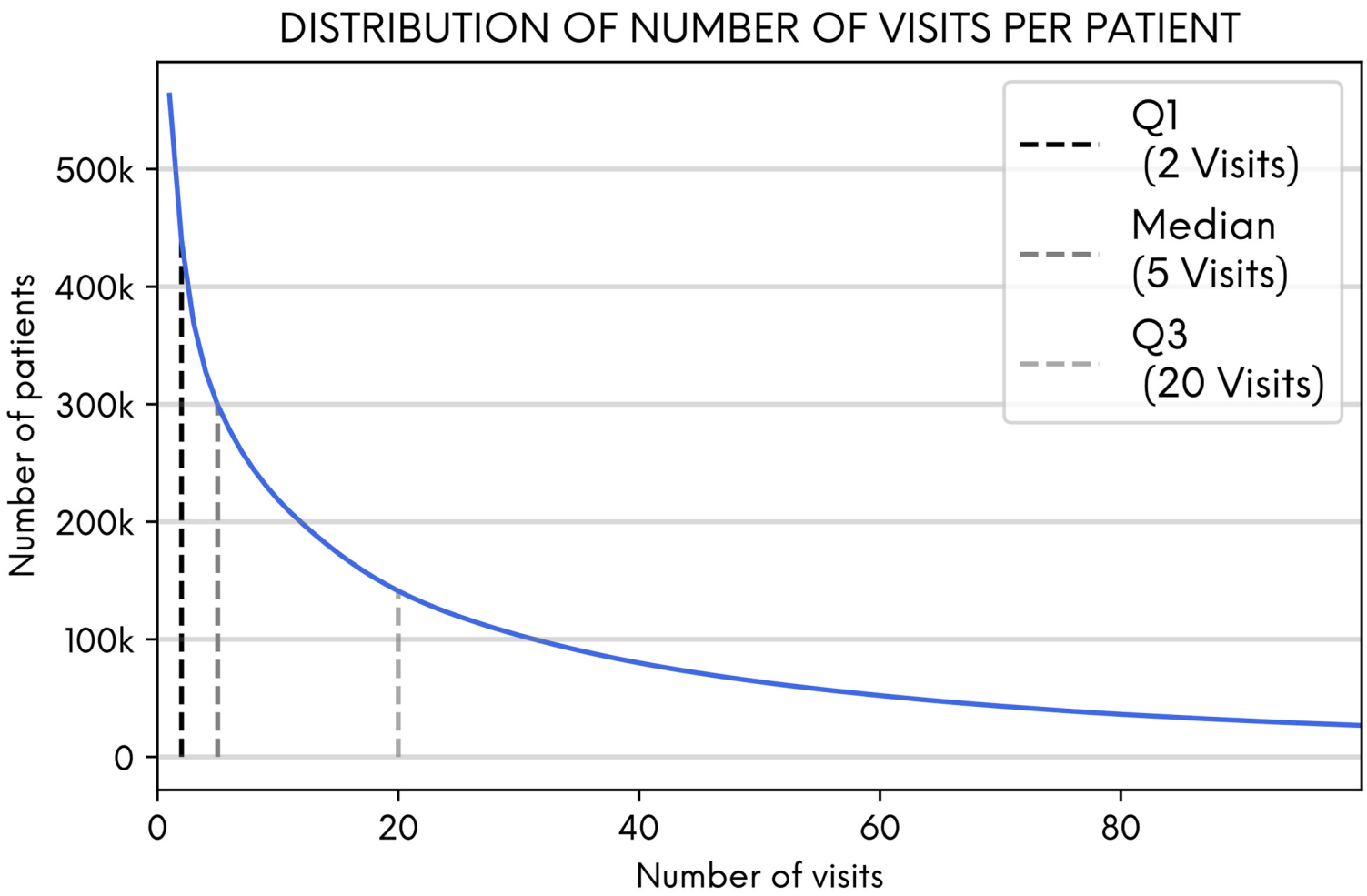
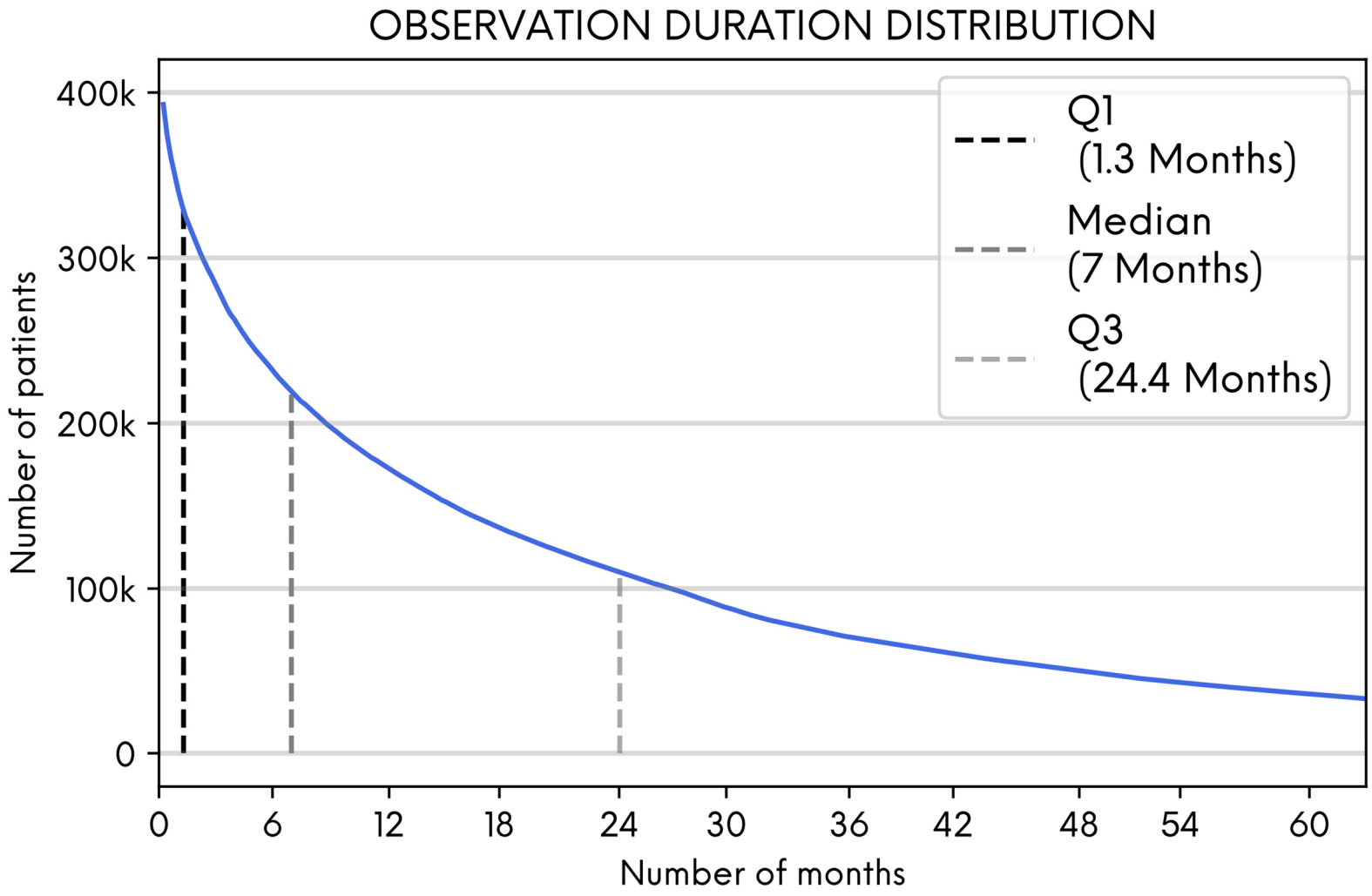


Online supplementary material

eFigure 10: Percentage of patients with at least one hospital visit by diagnosis



eFigure 11: Distribution of number of visits per patient



eFigure 12: Distribution of duration of follow-up